

Scoring table:	BLOSUM62	Query:	LKYAQICYWIKNYLDREG	Description
Searched:	GapOp 10.0 , capext 0.5	Match	383	
Total number of hits satisfying chosen parameters:	100059	Length	36664827 residues	
Minimum DB seq length: 0		ID		
Maximum DB seq length: 2000000000		DB		
Post-processing: Minimum Match 0%		ID		
Maximum Match 100%		DB		
Listing first 45 summaries		ID		
Database :	Swissprot_39*			
Pred.	No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.			
SUMMARIES				
Result No.	Score	Query	Match Length	DB ID
1	194	9.3	316	GSEP_BACLI
2	135.5	6.5	313	MPL_BACSU
3	124	6.0	266	ELI_BOVIN
4	121.5	5.8	490	FA0_RABIT
5	119	5.7	269	EL2_PIG
6	115.5	5.6	269	EL2_HUMAN
7	115	5.6	271	EL2_MOUSE
8	115	5.5	376	FA0_TROCA
9	114	5.5	238	TR5_AEDEA
10	113	5.4	266	EL1_PIG
11	113	5.4	266	EL1_RAT
12	112.5	5.4	259	TRP_STGR
13	112.5	5.4	273	YGD_ECOLI
14	111.5	5.4	488	FA0_HUMAN
15	111.5	5.4	875	NEPR_HUMAN
16	111	5.3	253	CFAD_HUMAN
17	111	5.3	761	NEPR_MOUSE
18	110	5.3	269	ELB_HUMAN
19	110	5.3	492	FA0_BOVIN
20	108	5.2	583	CFAL_HUMAN
21	108	5.2	786	STUB_DRONE
22	107.5	5.2	686	MAS2_HUMAN
23	107	5.1	603	CFAL_MOUSE
24	105.5	5.1	604	CFAL_RAT
25	105	5.0	269	EL2_BOVIN
26	105	5.0	274	TRV5_ANGCA
27	104	5.0	268	TRP_STGR
28	103.5	5.0	271	EL2_RAT
29	103.5	5.0	416	FA0_BOVIN
30	101.5	4.9	281	TRZ_DRONE
31	99	4.8	855	ST14_MOUSE
32	98.5	4.7	260	COGS_HYPLI
33	98.5	4.7	45	TRY7_ANGCA

ALIGNMENTS					
RESULT	1	GSEP_BACLI	STANDARD;	PRT;	316 AA.
ID	GSEP_BACLI				
AC	P80057;				
DT	01-MAY-1992 (Rel. 22, Created)				
DT	01-OCT-1993 (Rel. 27, Last sequence update)				
DT	20-AUG-2001 (Rel. 40, Last annotation update)				
DE	GLUTANYL ENDopeptidase PRECURSOR (EC 3.4.21.19) (GLUTAMATE SPECIFIC ENDOPEPTIDASE) (GSE).				
GN	BLASE.				
OS	Bacillus licheniformis.				
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillus.				
OX	NCBI_TAXID=1402;				
RN	[1]				
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.				
RC	STRAIN=ATCC 14580.				
RX	Medline=93054737; PubMed=1429718;				
RA	Kakudo S., Kikuchi N., Kitadokoro K., Fujiwara T., Nakamura E., Okamoto H., Shin M., Tamaki M., Terooka H., Tsuzuki H., Yoshiida N.;				
RT	"Purification, characterization, cloning, and expression of a glutamyl acid-specific protease from <i>Bacillus licheniformis</i> ATCC 14580."				
RT	J. Biol. Chem. 267:23782-23788(1992).				
RL	RN [2]				
RP	SEQUENCE 95%316.				
RX	Medline=92155199; PubMed=1346764;				
RA	Svendsen I., Breddam K.; "Isolation and amino acid sequence of a glutamyl acid specific endopeptidase from <i>Bacillus licheniformis</i> ."; Eur. J. Biochem. 204:165-171(1992).				
RT	"FUNCTION: SPECIFIC FOR HYDROLYSIS OF PEPTIDES BONDS ON THE CARBOXYL SIDE OF ACIDIC AMINO ACID RESIDUES, WITH A STRONG PREFERENCE FOR GLU. CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE: ASP-(-XAA, GLU-(-XAA). SUBCELLULAR LOCATION: SECRETED. SIMILARITY: BELONGS TO PEPTIDASE FAMILY 32B; ALSO KNOWN AS THE V8 FAMILY."				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).				
CC	EMBL: D10060; BAA00949.1; -.				
DR	PIR: S23078; S23078.				
DR	A45134; A45134.				
DR	MEROPS: S01_271; -.				
DR	InterPro: IPR00126; Ser_proteas_v8.				
DR	InterPro: IPR001254; Trypsin.				
DR	Pfam: PF00089; trypsin; 1.				
DR	PRINTS: PR00839; V8PROTEASE.				
DR	SMART: SM00020; tryp_spC; 1.				

PROSITE; PS00672; V8_HIS; 1.	RX MEDLINE-89108019; PubMed=3145906;
DR PROSITE; PS00673; V8_SER; 1.	RA Smith H., de Jong A., Bron S., Venema G.;
DR Hydrolase; Serine protease; Signal.	RT "Characterization of signal sequence-coding regions selected from the
SIGNAL 1 ?	RT Bacillus subtilis chromosome.";
SEQUENCE FROM N.A.	RT Gene 70:351-361(1988).
FT PROPEP ?	RL
FT CHAIN 95 316 GLUTAMYL ENDOPEPTIDASE.	RN [4]
FT ACT_SITE 141 141 CHARGE RELAY SYSTEM (BY SIMILARITY).	RN SEQUENCE OF 1-10 FROM N.A.
FT ACT_SITE 261 261 CHARGE RELAY SYSTEM (BY SIMILARITY).	RP STRAINE=168;
FT DISULFID 126 142	RC MEDLINE=96118702; PubMed=7496533;
FT DISULFID 275 279	RX SAXILD H.H., JACOBSEN J.H., NYGAARD P.;
SEQUENCE 316 AA: 33611 MW: 96D7552CB7089B09 CRC64;	RA "Functional analysis of the <i>Bacillus subtilis</i> purt gene encoding
	RT formate-dependent glycaminidase Ribonucleotide transformylase.";
Query Match 9.3%; Score 194; DB 1; Length 316;	RT Microbiology 141:221-228 (1995).
Best Local Similarity 24.0%; Pred. No. 4.3e-09;	RL
Matches 79; Conservative 43; Mismatches 143; Indels 64; Gaps 15;	CC -1 - SUBUNIT: MONOMER.
Qy 63 SSSGPQCHKGTFLL--PTYEERKQYLSYETLXANGSRTEQVGYIILSSSGGAGQRDSD 119	CC -1 - SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2B; ALSO KNOWN AS THE S2
Db 28 AQQAPSPH-TPVSSDPSY-KAETSVTYDP----NIKSDQYGLYSKAFTGTGVNETKE 79	CC -1 - CAUTION: CALLED "METALLOPROTASE", BUT CLEARLY BELONGS TO THE S2
Qy 120 SGSKSRRRKQYIYGDSRFSRKFDFELLN----YFESTSYKLST--GCYGTLYAEKHV 170	CC -1 - FAMILY OF SERINE PROTEASIS.
Db 80 KAEKSPAKARY--SIKSVKQDDTRTNTAYPAIWHISSGCTQWMICPKTV 136	CC
Qy 171 LTAACHTDGGKT-YVKGTQKLRLRGFLKPKEFDGGGRANDSTSAMPQMKFQWIRVKRTHY 229	CC This SWISS-PROT entry is copyright. It is produced through a collaboration
Db 137 ATAGHCYDTSGSFACTATPSG-----RNGTS-----RNQTS-----RNGTS 178	CC between the Swiss Institute of Bioinformatics and the EMBL outstation
Qy 230 PKGWIGNANDGMDDYDALLYELKKDKHKKMKWIGSPPARQCLPGGRHFSGYDNRPGN 289	CC the European Bioinformatics Institute. There are no restrictions on its
Db 179 PSQWRSCNTN----YDGATELSEPIGNTRYGFSYTTSSLYGTIVTISYPGDKTAG 233	CC use by non-profit institutions as long as its content is in no way
Qy 290 LYVRECD--VKDETDLYQODQAQPGAGSSCVYRMWKRQOKWERKII---GIFSGH 343	CC modified and this statement is not removed. Usage by and for commercial
Db 234 TQWQHSRPIASETYKQYAMDTYQGQSGSPVFEQSSSRNCGFCPSLAVHTNGVYGG 291	CC entities requires a license agreement. (See http://www.isb-sib.ch/announcecc/)
Qy 344 QWDMMNGSPQDENVAIRTPKLYAQICYW 372	CC or send an email to license@isb-sib.ch).
Db 292 -----SSINRGRITKEVFDNLTNW 311	CC
RESULT 2	CC
MR_BACSU STANDARD; PRT; 313 AA.	CC
ID MPR_BACSU	CC
AC P39790	CC
STRAIN 1423;	CC
SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.	CC
STRAIN GP241;	CC
MEDLINE=90130256; PubMed=2105291;	CC
ALLY D., RUDOLPH C.F., RUFO G.A. JR., SULLIVAN B.J., THERIAULT K.A.,	CC
"Gene encoding a novel extracellular metalloprotease in <i>Bacillus</i>	CC
subtilis";	CC
J. Bacteriol. 172:1024-1029(1990).	CC
OS Bacillus subtilis.	CC
OC Firmicutes; <i>Bacillus/Clostridium</i> group;	CC
OC <i>Bacillus/Steaphylococcus</i> group; <i>Bacillus</i> .	CC
OX NCBI_TAXID=1423;	CC
RN [1]	CC
RF 01-FEB-1995 (Rel. 31, Created)	CC
RF 01-FEB-1995 (Rel. 31, Last sequence update)	CC
RF 20-AUG-2001 (Rel. 40, Last annotation update)	CC
DE EXTRACELLULAR METALLOPROTEASE PRECURSOR (EC 3.4.21.-).	CC
GN MPR	CC
OS	CC
OC Bacteria; Firmicutes; <i>Bacillus/Clostridium</i> group;	CC
OC <i>Bacillus/Steaphylococcus</i> group; <i>Bacillus</i> .	CC
OX NCBI_TAXID=1423;	CC
RN [1]	CC
RF SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.	CC
RF SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.	CC
RF STRAIN GP241;	CC
RF MEDLINE=90130256; PubMed=2105291;	CC
RA SULLIVAN B.J., THERIAULT K.A.,	CC
RA ALY D., RUDOLPH C.F., RUFO G.A. JR., SULLIVAN B.J., THERIAULT K.A.,	CC
RA "Gene encoding a novel extracellular metalloprotease in <i>Bacillus</i>	CC
RA subtilis";	CC
RA J. Bacteriol. 172:1024-1029(1990).	CC
RN [2]	CC
SEQUENCE FROM N.A.	CC
STRAIN=168;	CC
RC	CC
RN 165 VASKHVLDAAHCHH-----DGKTYVKQTQKLVRGFPLKPKFDGGRND 208	CC
RA Haga K., Liu H., Yasumoto K., Takahashi H., Yoshikawa H.;	DB 136 VNPNVVAGHCYVQSDIGHWASTITAAPGRNGSSYYPYTY-----175
RT "Sequence analysis of the 70kb region between 17 and 23 degree of the	RT Bacillus subtilis chromosome.";
RT Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.	RT
RP SEQUENCE OF 1-68 FROM N.A.	RP

Result No.	Score	Query Match	Length	DB ID	Description
1	194	9.3	316	2 A45134	endopeptidase (EC metalloproteinase
2	135.5	6.5	313	2 A35122	coagulation factor
3	120.5	5.8	482	1 EXRT	pancreatic elastase
4	119	5.7	269	2 A26823	pancreatic elastase
5	117	5.6	522	2 T29167	pancreatic elastase
6	115.5	5.6	269	2 B26823	pancreatic elastase
7	115	5.6	271	2 A25528	pancreatic elastase
8	115	5.5	258	2 S70439	probable pancreatic elastase
9	115	5.5	267	4 A56615	trypsin-like protease
10	114.5	5.5	238	1 TRNV5Y	coagulation factor
11	114	5.5	246	1 DBHU	pancreatic elastase
12	113	5.4	266	1 ELPT1	pancreatic elastase
13	113	5.4	266	1 ELPG	trypsin (EC 3.4.21)
14	111.5	5.4	259	1 TRSG	hypothetical protease
15	112	5.4	273	2 E85765	probable secreted complement factor
16	112.5	5.4	273	2 H64915	hypothetical protease
17	111.5	5.4	488	1 EXHU	coagulation factor
18	111	5.3	761	2 JC3759	brain-specific serine proteinase
19	110	5.3	269	2 C26823	pancreatic elastase
20	110	5.3	492	1 EXBO	coagulation factor
21	110	5.3	1582	2 T13108	hypothetical protease
22	109.5	5.3	405	2 T35117	probable secreted complement factor
23	108	5.2	583	2 A29154	serine proteinase
24	108	5.2	786	1 A41547	masquerade precursors
25	108	5.2	1047	2 A55617	T-cell suppressor
26	107.5	5.2	236	2 A25566	T-cell reactive factor
27	107	5.2	686	1 A5971	trypsin-related proteases
28	105	5.0	274	2 SA4004	pancreatic elastase
29	103.5	5.0	271	1 ELRT2	

Copyright (c) 1993 - 2000 Compugen Ltd.
 OM protein - protein search, using sw model
 Run on: December 2, 2001, 16:17:56 ; Search time 48.03 Seconds
 (without alignments)
 607.430 Million cell updates/sec

Title: US-09-072-384-18
 Perfect score: 2080
 Sequence: 1 MAGIPGLLFLLELLCAVQGQ.....LKYAQICYWIKGNYLDREG 383

Scoring table: BL0SUM62
 Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
 Maximum Match 100%
 Listing first 45 summaries

Database : PIR_68:
 1: pir1:
 2: pir2:
 3: pir3:
 4: pir4:
 5: pir5:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	No.	Score	Query Match	Length	DB ID	Description	Best Local Similarity	Score	DB 2;	Length
1	1	194	9.3	316	2 A45134	endopeptidase (EC metalloproteinase	24.0%	24.0%	DB 2;	Length 316;
2	2	135.5	6.5	313	2 A35122	coagulation factor	24.0%	24.0%	DB 2;	Length 316;
3	3	120.5	5.8	482	1 EXRT	pancreatic elastase	24.0%	24.0%	DB 2;	Length 316;
4	4	119	5.7	269	2 A26823	pancreatic elastase	24.0%	24.0%	DB 2;	Length 316;
5	5	117	5.6	522	2 T29167	pancreatic elastase	24.0%	24.0%	DB 2;	Length 316;
6	6	115.5	5.6	269	2 B26823	pancreatic elastase	24.0%	24.0%	DB 2;	Length 316;
7	7	115	5.6	271	2 A25528	pancreatic elastase	24.0%	24.0%	DB 2;	Length 316;
8	8	115	5.5	258	2 S70439	probable pancreatic elastase	24.0%	24.0%	DB 2;	Length 316;
9	9	115	5.5	267	4 A56615	trypsin-like protease	24.0%	24.0%	DB 2;	Length 316;
10	10	114.5	5.5	238	1 TRNV5Y	coagulation factor	24.0%	24.0%	DB 2;	Length 316;
11	11	114	5.5	246	1 DBHU	pancreatic elastase	24.0%	24.0%	DB 2;	Length 316;
12	12	113	5.4	266	1 ELPT1	pancreatic elastase	24.0%	24.0%	DB 2;	Length 316;
13	13	113	5.4	266	1 ELPG	trypsin (EC 3.4.21)	24.0%	24.0%	DB 2;	Length 316;
14	14	111.5	5.4	259	1 TRSG	hypothetical protease	24.0%	24.0%	DB 2;	Length 316;
15	15	112	5.4	273	2 E85765	probable secreted complement factor	24.0%	24.0%	DB 2;	Length 316;
16	16	112.5	5.4	273	2 H64915	hypothetical protease	24.0%	24.0%	DB 2;	Length 316;
17	17	111.5	5.4	488	1 EXHU	coagulation factor	24.0%	24.0%	DB 2;	Length 316;
18	18	111	5.3	761	2 JC3759	brain-specific serine proteinase	24.0%	24.0%	DB 2;	Length 316;
19	19	110	5.3	269	2 C26823	pancreatic elastase	24.0%	24.0%	DB 2;	Length 316;
20	20	110	5.3	492	1 EXBO	coagulation factor	24.0%	24.0%	DB 2;	Length 316;
21	21	110	5.3	1582	2 T13108	hypothetical protease	24.0%	24.0%	DB 2;	Length 316;
22	22	109.5	5.3	405	2 T35117	probable secreted complement factor	24.0%	24.0%	DB 2;	Length 316;
23	23	108	5.2	583	2 A29154	serine proteinase	24.0%	24.0%	DB 2;	Length 316;
24	24	108	5.2	786	1 A41547	masquerade precursors	24.0%	24.0%	DB 2;	Length 316;
25	25	108	5.2	1047	2 A55617	T-cell suppressor	24.0%	24.0%	DB 2;	Length 316;
26	26	107.5	5.2	236	2 A25566	T-cell reactive factor	24.0%	24.0%	DB 2;	Length 316;
27	27	107	5.2	686	1 A5971	trypsin-related proteases	24.0%	24.0%	DB 2;	Length 316;
28	28	105	5.0	274	2 SA4004	pancreatic elastase	24.0%	24.0%	DB 2;	Length 316;
29	29	103.5	5.0	271	1 ELRT2		24.0%	24.0%	DB 2;	Length 316;

coagulation factor
 hypothetical protease
 trypsin (EC 3.4.21)
 probable trypsin V
 probable polygalactose
 probable secreted
 coagulation factor
 complement factor
 coagulation factor
 complement factor
 allergen Der f III
 factor IX - rabbit
 adenylyl cyclase
 serine proteinase
 hypothetical protease
 pancreatic elastase

RESULTS

ALIGNMENTS

coagulation factor
 glutamate-specific - *Bacillus licheniformis*
 endopeptidase (EC 3.4.21), glutamate-specific - *Bacillus licheniformis*
 C:Species: *Bacillus licheniformis*
 C:Date: 10-Jun-1993 #sequence_revision 18-Nov-1994 #text_change 15-Oct-1999
 C:Accession: A45134; S23078
 R:Kakudo, S.; Kiuchi, N.; Kitadokoro, K.; Fujiwara, T.; Nakamura, E.; Okamoto, H.; S. J. Biol. Chem. 267, 23782-23788, 1992
 A:Title: Purification, characterization, cloning, and expression of a glutamic acid specific endopeptidase
 A:Reference number: A45134; MUID:93034737
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-316 <KAK>
 A:Cross-references: GB:D10060; NID:9216263; PID:BA000949.1; PID:d1001415; PID:g21626
 A:Experimental source: ATCC 14580
 A:Note: sequence extracted from NCBI backbone (NCBIN:118784, NCBIPI:118785)
 R:Svendsen, I.; Breddam, K.
 Eur. J. Biochem. 204, 165-171, 1992
 A:Title: Isolation and amino acid sequence of a glutamic acid specific endopeptidase
 A:Reference number: S23078
 A:Status: preliminary
 A:Molecule type: protein
 A:Keywords: hydrolase

Query Match 9.38; Score 194; DB 2; Length 316;
 Best Local Similarity 24.0%; Pred. No. 1.2e-08; Mismatches 143; Indels 64; Gaps 15;

QY 63 SSGCPQCHKGTPL--PTYEAKQYLSEKTYANGSRTETQGYIILSSSGDGAQRHDSG 119
 Db 28 AQAAQSPH--TPEVSSDPSY-KAETSVTYD----NKSQDGLYSRAFTGKVNETKE 79

Query Match 9.38; Score 194; DB 2; Length 316;
 Best Local Similarity 24.0%; Pred. No. 1.2e-08; Mismatches 143; Indels 64; Gaps 15;

QY 120 SSGCKSRKRQTYGDSRSTIFGRDLFLN--YVPSVTSVGLVWVSSSSGCTGWMIGKTV 170
 Db 80 KAEKKSPAKAPY--SIKSVIGSDDPRVTNTAYPRAVHIVSSSSGCTGWMIGKTV 136

QY 171 LTAAHCTHDGKT-YVKGTQKLRYGVFLPKPFKDGGRGANDSTSAMPQQMKFOWIRVKRTHV 229

Db 137 AVAGHCYDTSSCSFACTAVSPG-----RNGTS-----YPYGEVKSTRYFI 178

QY 230 PKGWIGKANDIGMDYALLELKKPHRKEMKIGVSSPAKOLPGGRIFHSGYDNDPQN 289

Db 179 PSGWRSNTN----YOGAIESELPEIGNTVGFLSYTTSLVGTIVTSCYPGDRTAG 233

QY 290 LYVRFCD--VKBDTYDLYQQCDAQPSGASGSYVVRMKRQOKWEEKII---GIFSGH 343

Db 234 TQWQHSGSPIAISETYKLQYAM-DTYGGQSNSPWEQSSSRNTCSGPCLSLAVHTNGYYGG- 291

QY 344 QWDMNGSPQDENYAVRITPLKYAQICYW 372
Db 292 -----SSYNRGTTRTKEFNLTNW 311

RESULT 2
A35122 metalloproteinase (EC 3.4.-.-) mpr precursor, extracellular - *Bacillus subtilis*
C;Species: *Bacillus subtilis*
C;Date: 27-Jul-1990 #sequence_revision 27-Jul-1990 #text_change 15-Oct-1999
R;Sloma, A.; Rudolph, C.F.; Rufo, Jr., G.A.; Sullivan, B.J.; Theriault, K.A.; Ally, D.; R.; Smith, H.; de Jong, A.; Bron, S.; Verema, G.
A;Title: Gene encoding a novel extracellular metalloprotease in *Bacillus subtilis*.
A;Reference number: A35122; MUID:00130256
A;Accession: A35122
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-313 <SL0>
A;Cross-references: GB:L10505; PIDN:AAA22604.1; PID:g143210; GB:M29036
A;Reference number: 139594; MUID:89108019
A;Title: Characterization of signal-sequence-coding regions selected from the *Bacillus s*
A;Accession: I40010
A;Status: preliminary; translated from GB/EMBL/DDJB
A;Molecule type: DNA
A;Residues: 1-60,65,'L',67,'S',69,'AQ' <RES>
A;Cross-references: GB:M22916; PID:g143702
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azavedo, V.; Bertani, C.; Bron, S.; Brussel, S.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choc, A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 320, 249-256, 1997
A;Authors: Foulier, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallerie, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullio, M.F.; Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, F.; Lauber, J.; Lazarovic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mauell, Y.; Ogawa, K.; Ogiwara, A.; Onida, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, D.; Rieger, M.; Rivolta, C.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon, A; Authors: Schleich, S.; Schrofe, R.; Scrofe, P.; Terpstra, P.; Tognoni, A.; Tosato, V.; Seror, Akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Uchiyama, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.; Authors: Yoshioka, H.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.
A;Reference number: A69380; MUID:98044033
A;Accession: A69660
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-313 <KUN>
A;Cross-references: GB:AL009126; PID:g2632457; PID:CAB12018.1; PID:e1182176;
A;Experimental source: strain 168
C;Genetics:
A;Gene: mpr
C;Keywords: hydrolase

Query Match 6.5% Score 135.5; DB 2; Length 313;
Best Local Similarity 22.2%; Pred. No. 0.00091; Matches 38; Mismatches 121; Indels 111; Gaps 17;

QY 72 GTPPIPYERAKQYLSYELYANGSRTETQVGIYLSSLGGD-GAQHRDSSGSKSRKKRQ 129
Db 29 GVPKAAGNPQTSNTGKEADATKNOT-----SKADQVSAPYESTGKTSKS----- 75
QY 130 IGYDSRF-----SIFGKD-----FLINYPFESTVSKLST-----GCMTGTL 164
Db 76 LYGGTQELEKNQTQLPQSSITGDERTRISSTSFPPYRQLSIKYKPNTSSTYGCGLF 135
QY 165 VAERKHVLTAAHCIIH-----DGKTYVKGTOKLRVGFKLKPKFKDGGGRAND 208
Db 136 VNPPTVTVAGCHYQSQDHGWAStITAAPGRNGSSPYTY----- 175
QY 209 STSAMPEQMKEFWIRVKRTHPVKGWIKGN--ANDIGMDYDVALLELKPKHFKMFKIGV 265

Db 176 -SGTMFYSVK-GWTESKDNYDGAIKLNGSPGNTVGW-YGYRTNSSLSS-----VGL 225
QY 266 SPPAKOLPGGRHSGYDNDRPGNLVYRCQCDADPASGSYYVRMW 325
Db 226 SSSVTGFPCCDKTFCTMMWSDTKPIR-----SAETYKLY-TDITYGCOSGPVY--- 272
QY 326 KRQQQKWERLIGFGHWDMMGSPQDENAVAYTRPLKYAQICYW 372
Db 273 -RNYSDTGQTAIHT-----NGG-SSYNLGTRVTDVFNNIQYW 310

RESULT 3
EXRT coagulation factor Xa (EC 3.4.21.6) precursor - rat
C;Species: *Rattus norvegicus* (Norway rat)
C;Accession: S49075; PID:9506600; PIDN:CAA56202.1; PID:9506601
R;Stanton, C.; Ross, P.; Huston, S.; Wallin, R.
Thromb. Res. 80, 63-73, 1995
A;Title: Evidence for competition between vitamin K-dependent clotting factors
A;Reference number: A58498; MUID:36093366
A;Accession: S49075
A;Molecule type: mRNA
A;Residues: 1-482 <STAL>
A;Cross-references: EMBL:X79807; PID:9506600; PIDN:CAA56202.1; PID:9506601
A;Note: Submitted to the EMBL Data Library, June 1994
A;Note: neither the complete nucleic acid sequence nor the complete translation are s
R;Stanton, C.; Ross, P.; Huston, S.; Wallin, R.
Gene 169, 269-273, 1996
A;Title: Processing and expression of rat and human clotting factor-X-encoding cDNAs.
A;Reference number: JC4670; MUID:96194815
A;Accession: JC4670
A;Molecule type: mRNA
A;Residues: 1-482 <STAL>
A;Cross-references: EMBL:X79807; PID:9506600; PIDN:CAA56202.1; PID:9506601
A;Experimental source: Cos-1 cell
R;Ejyoji, K.; Mizazaki, K.; Kata, H.
J. Biochem. 109, 890-898, 1991
A;Title: Characterization of rat factors X and Xa: demonstration of factor Xa in rat
A;Reference number: PS0190; MUID:32041742
A;Accession: PS0191
A;Molecule type: protein
A;Residues: 41-58 /X/ 60-65 <ENJ1>
A;Accession: PS0190
A;Molecule type: protein
A;Residues: 1-482 <STAL>
A;Cross-references: EMBL:X79807; PID:9506600; PIDN:CAA56202.1; PID:9506601
A;Molecule type: DNA
A;Residues: 183-186 /X/ 188-207 <ENJ2>
R;Murakawa, M.; Okamura, T.; Kamura, T.; Kuroiwa, M.; Niho, Y.
Eur. J. Haematol. 52, 162-168, 1994
A;Title: Analysis of the partial nucleotide sequences and deduced primary structure
A;Reference number: I46196; MUID:94222160
A;Accession: I46196
A;Function: catalyzes the proteolytic activation of prothrombin to thrombin in the
A;Pathway: blood coagulation
C;Superfamily: coagulation factor X; EGF homology; Glu domain homology; trypsin homolog
C;Keywords: beta-hydroxyaspartic acid; blood coagulation; calcium binding; carboxyglu
F;1-23/Domain: signal sequence #status predicted <SIG>
F;24-40/Domain: propeptide #status predicted <PRO>
F;25-84/Domain: Glu domain homology <GLA>
F;41-179/Product: coagulation factor X light chain #status predicted <LCH>
F;90-121/Domain: EGF homology <EGI>
F;129-164/Domain: EGF homology <EG2>
F;183-482/Product: coagulation factor X heavy chain #status predicted <HCH>
F;183-231/Domain: activation peptide #status predicted <AP1>
F;22-460/Domain: coagulation factor Xa heavy chain #status predicted <RCT>
F;22-460/Domain: trypsin homology <TRY>
F;46-47 54, 56, 59 60, 65, 66, 69, 72, 79/Modified site: gamma-carboxyglutamic acid (Glu) #S
F;57-62, 90-101, 95-110, 112-121, 129-140, 136-149, 151-164, 172-340, 238-243, 259-275, 388-402

Result No.	Score	Query	Match	Length	DB ID	Description
1	2080	97.6	383	4	Q95084	095084 homo sapien
2	1884.5	88.4	382	11	Q9d6X6	Q9d6X6 mus musculus
3	1042	48.9	413	4	Q9BQP6	Q9BQP6 homo sapien
4	129.5	6.3	799	11	Q9DBT0	Q9DBT0 mus musculus
5	129.5	6.1	303	2	Q9EXR9	Q9EXR9 bacillus in
6	127.9	6.1	469	6	Q9GMD9	Q9GMD9 ornithorhynchus anatinus
7	127.5	6.0	1322	11	Q9NATO	Q9NATO anophelles gambiae
8	126.5	5.9	678	11	Q9JJS8	Q9JJS8 rattus norvegicus
9	124.5	5.8	1322	5	Q9NJS5	Q9NJS5 anophelles gambiae
10	124	5.8	266	6	Q9644	Q9644 macaca fasciata
11	123.5	5.8	339	11	Q9QX91	Q9QX91 rattus norvegicus
12	123.5	5.8	366	11	Q9QX85	Q9QX85 rattus norvegicus
13	123.5	5.8	541	11	Q9QX90	Q9QX90 rattus norvegicus
14	123.5	5.8	623	11	Q9JJP3	Q9JJP3 rattus norvegicus
15	123.5	5.8	643	11	Q9QX84	Q9QX84 rattus norvegicus
16	122	5.7	259	5	Q9XY61	Q9XY61 ctenocephalid larva
17	122	5.7	449	5	Q9VYD8	Q9VYD8 drosophila melanogaster
18	120.5	5.7	482	11	Q63207	Q63207 rattus norvegicus
19	118	5.5	266	11	Q9d936	Q9d936 mus musculus

Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2001, 16:25:48 ; Search time 84.84 Seconds
(without alignments)

675.846 Million cell updates/sec

Title: US-09-072-384-15

Perfect score: 2131

Sequence: 1 MAGIPGLLFLLEFLCAVGQ IKGNYLDCREGDTVFPFGSN 392

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL_17:
1: sp_archaea:
2: sp_bacteria:
3: sp_fungi:
4: sp_human:
5: sp_invertebrate:
6: sp_mammal:
7: sp_mhc:
8: sp_organelle:
9: sp_phage:
10: sp_plant:
11: sp_rabbit:
12: sp_virus:
13: sp_vertebrate:
14: sp_unclassified:
15: sp_worm:

ALIGNMENTS

RESULT 1
095084 PRELIMINARY; PRT: 383 AA.
ID 095084,
AC 095084,
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE SERINE PROTEASE (HYPOTHETICAL 4.0 KDA PROTEIN) (PROTEASE, SERINE, 23).
DE 23).
GN 2SIG13 OR DKFZP586B0719.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. NCBITaxid=9606;
OX RN [1]
RN SEQUENCE FROM N.A.
RN TISSUE=UMBILICAL VEIN;
RC Sheppard P., Blumberg H., Jelinek L., Foster D., O'Hara P.; Li X., Tedder T.F.; RT "A novel serine protease from human umbilical vein endothelial cells." RLT Submitted (JUL-1997) to the EMBL/GenBank/DDBJ databases.
RN [2]
RN SEQUENCE FROM N.A.
RP RA Sheppard P., Blumberg H., Jelinek L., Foster D., O'Hara P.; RA Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
RN [3]
RN SEQUENCE FROM N.A.
RP RA Wambutt R., Heubner D., Newes H.W., Gassenhuber J., Wiemann S.; RA Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN [4]
RN SEQUENCE FROM N.A.
RC Strausberg R.; RA Submitted (DEC-2000) to the EMBL/GenBank/DDBJ databases.
RL CC -1- SIMILARITY: TO CHYMOTRYPSIN SERINE PROTEASE FAMILY (S1). DR EMBL; AF015287; AAD01553.1; DR EMBL; AF193611; AAF07186.1; DR EMBL; AL136914; CAB65848.1; DR EMBL; BC001278; AAH01278.1; DR MEROPS; S01.309; S01.309;

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Match	Length	DB ID	Description
1	2080	97.6	383	4	Q95084	095084 homo sapien
2	1884.5	88.4	382	11	Q9d6X6	Q9d6X6 mus musculus
3	1042	48.9	413	4	Q9BQP6	Q9BQP6 homo sapien
4	129.5	6.3	799	11	Q9DBT0	Q9DBT0 mus musculus
5	129.5	6.1	303	2	Q9EXR9	Q9EXR9 bacillus in
6	127.9	6.1	469	6	Q9GMD9	Q9GMD9 ornithorhynchus anatinus
7	127.5	6.0	1322	11	Q9NATO	Q9NATO anophelles gambiae
8	126.5	5.9	678	11	Q9JJS8	Q9JJS8 rattus norvegicus
9	124.5	5.8	1322	5	Q9NJS5	Q9NJS5 anophelles gambiae
10	124	5.8	266	6	Q9644	Q9644 macaca fasciata
11	123.5	5.8	339	11	Q9QX91	Q9QX91 rattus norvegicus
12	123.5	5.8	366	11	Q9QX85	Q9QX85 rattus norvegicus
13	123.5	5.8	541	11	Q9QX90	Q9QX90 rattus norvegicus
14	123.5	5.8	623	11	Q9JJP3	Q9JJP3 rattus norvegicus
15	123.5	5.8	643	11	Q9QX84	Q9QX84 rattus norvegicus
16	122	5.7	259	5	Q9XY61	Q9XY61 ctenocephalid larva
17	122	5.7	449	5	Q9VYD8	Q9VYD8 drosophila melanogaster
18	120.5	5.7	482	11	Q63207	Q63207 rattus norvegicus
19	118	5.5	266	11	Q9d936	Q9d936 mus musculus

DR	InterPro; IPR001314; Chymotrypsin.	RA	Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N., Saeki H., Sato K., Schoenbach C., Storch K.-F., Storch Y., Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L., Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S., Hayashizaki Y.; "Functional annotation of a full-length mouse cDNA collection.";
DR	Pfam; PF00089; trypsin; 1.	RA	"Functional annotation of a full-length mouse cDNA collection.";
DR	PRTS; PR00722; CHYMOTRYPSIN.	RA	RT
DR	SMART; SM00202; TRYPSIN_SPC; 1.	RA	"Functional annotation of a full-length mouse cDNA collection.";
DR	PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.	RA	RT
KW	Hydrolase; Protease; Serine protease.	RA	Nature 409:695-690(2001).
SEQUENCE	383 AA; 43001 MW; 4666C11ABFD5E8F CRC64;	RL	-1 - SIMILARITY: TO CHYMOTRYPSIN SERINE PROTEASE FAMILY (S1).
CC	CC -1 - SIMILARITY: TO CHYMOTRYPSIN SERINE PROTEASE FAMILY (S1).	CC	CC -1 - SIMILARITY: TO CHYMOTRYPSIN SERINE PROTEASE FAMILY (S1).
DR	EMBL; AK009847; BAB26501.1; -	DR	EMBL; AK009847; BAB26501.1; -
DR	MGT; MG1:19237/03; 231046G15Rik.	DR	MGT; MG1:19237/03; 231046G15Rik.
DR	InterPro; IPR001314; Chymotrypsin.	DR	InterPro; IPR001314; Chymotrypsin.
DR	InterPro; IPR001254; Trypsin.	DR	InterPro; IPR001254; Trypsin.
PFAM	PF00088; trypsin; 1.	DR	PFAM; PF00088; trypsin; 1.
PRINTS	PRO0722; CHYMOTRYPSIN.	DR	PRINTS; PRO0722; CHYMOTRYPSIN.
SMART	SM00020; TRYSPC; 1.	DR	SMART; SM00020; TRYSPC; 1.
PROSITE	PS00134; TRYSPIN_HIS; UNKNOWN_1.	KW	PROSITE; PS00134; TRYSPIN_HIS; UNKNOWN_1.
HYDROLASE	Serine protease.	KW	HYDROLASE; Serine protease.
SEQUENCE	382 AA; 43053 MW; 69C9A7080E5B2306 CRC64;	SQ	SEQUENCE
Query	Match 97.6%; Score 2080; DB 4; Length 383;	Query	Match 88.4%; Score 1884.5; DB 11; Length 382;
Best Local Similarity 100.0%; Pred. No. 1.8e-182;	Best Local Similarity 90.6%; Pred. No. 1.e-164;	Best Local Similarity 90.6%; Pred. No. 1.e-164;	Best Local Similarity 90.6%; Pred. No. 1.e-164;
Matches 383; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	Matches 347; Conservative 15; Mismatches 20; Indels 1; Gaps 1;	Matches 347; Conservative 15; Mismatches 20; Indels 1; Gaps 1;	Matches 347; Conservative 15; Mismatches 20; Indels 1; Gaps 1;
OY	1 MAGIPGLLFLFLFLCLAVQVSPYSAPKPPTPAYRLPVLPVILPOSTLNIAKPDFGAEAKLE 60	OY	1 MAGIPGLLFLFLFLCLAVQVSPYSAPKPPTPAYRLPVLPVILPOSTLNIAKPDFGAEAKLE 60
Db	1 MAGIPGLLFLFLFLCLAVQVSPYSAPKPPTPAYRLPVLPVILPOSTLNIAKPDFGAEAKLE 60	Db	1 MAGIPGLLFLFLFLCLAVQVSPYSAPKPPTPAYRLPVLPVILPOSTLNIAKPDFGAEAKLE 60
OY	61 VSSSGPQCHKGTPPLPTRYEAQKOLSYSTLYANGSRTRTOQYIYTLLSSGDGAQHRSGS 120	OY	61 VSSSGPQCHKGTPPLPTRYEAQKOLSYSTLYANGSRTRTOQYIYTLLSSGDGAQHRSGS 120
Db	61 VSSSGPQCHKGTPPLPTRYEAQKOLSYSTLYANGSRTRTOQYIYTLLSSGDGAQHRSGS 120	Db	61 VSSSGPQCHKGTPPLPTRYEAQKOLSYSTLYANGSRTRTOQYIYTLLSSGDGAQHRSGS 120
OY	121 SGSSRKKRQIYGDSRFSTFGKDFLNNPFSTSVKLSTRGCTGTVAAKHLTAACHTHDG 180	OY	121 SGSSRKKRQIYGDSRFSTFGKDFLNNPFSTSVKLSTRGCTGTVAAKHLTAACHTHDG 180
Db	121 SGSSRKKRQIYGDSRFSTFGKDFLNNPFSTSVKLSTRGCTGTVAAKHLTAACHTHDG 180	Db	121 SGSSRKKRQIYGDSRFSTFGKDFLNNPFSTSVKLSTRGCTGTVAAKHLTAACHTHDG 180
OY	181 KTVVKGTOKLKVGLPKKFQKDFGRGRANDSTSAMPQMFOWIRVKRTHVPKGWIKGNAND 240	OY	181 KTVVKGTOKLKVGLPKKFQKDFGRGRANDSTSAMPQMFOWIRVKRTHVPKGWIKGNAND 240
Db	181 KTVVKGTOKLKVGLPKKFQKDFGRGRANDSTSAMPQMFOWIRVKRTHVPKGWIKGNAND 240	Db	181 KTVVKGTOKLKVGLPKKFQKDFGRGRANDSTSAMPQMFOWIRVKRTHVPKGWIKGNAND 240
OY	241 IGMDYDALLEKKPHKRKFMKIGVSPPAKOLPGGRHIFSGYDNDRPGNLVYRFCDYKDE 300	OY	241 IGMDYDALLEKKPHKRKFMKIGVSPPAKOLPGGRHIFSGYDNDRPGNLVYRFCDYKDE 300
Db	241 IGMDYDALLEKKPHKRKFMKIGVSPPAKOLPGGRHIFSGYDNDRPGNLVYRFCDYKDE 300	Db	241 IGMDYDALLEKKPHKRKFMKIGVSPPAKOLPGGRHIFSGYDNDRPGNLVYRFCDYKDE 300
OY	301 TYDLYQQCDAPQGASGSVYVMWKROQQKWERKIGIFSGHONVDMNGSPQDENYAVR 360	OY	301 TYDLYQQCDAPQGASGSVYVMWKROQQKWERKIGIFSGHONVDMNGSPQDENYAVR 360
Db	301 TYDLYQQCDAPQGASGSVYVMWKROQQKWERKIGIFSGHONVDMNGSPQDENYAVR 360	Db	301 TYDLYQQCDAPQGASGSVYVMWKROQQKWERKIGIFSGHONVDMNGSPQDENYAVR 360
OY	361 ITPLKYAQICYWIKGNYLDCREG 383	OY	361 ITPLKYAQICYWIKGNYLDCREG 383
Db	361 ITPLKYAQICYWIKGNYLDCREG 383	Db	361 ITPLKYAQICYWIKGNYLDCREG 383
RESULT 2 Q9DX6	PRELIMINARY; PRINT; 382 AA.	RESULT 3 Q9BOP6	PRELIMINARY; PRINT; 413 AA.
ID	Q9DX6	ID	Q9BOP6
AC	Q9DX6_X6	AC	Q9BOP6
DT	01-JUN-2001 (TREMBLrel. 17, Created)	DT	01-JUN-2001 (TREMBLrel. 17, Created)
DT	01-JUN-2001 (TREMBLrel. 17, Last sequence update)	DT	01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)	DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE	2310046G15RIK PROTEIN.	DE	DJ223E3.1 (PUTATIVE SECRETED PROTEIN ZSIG13).
GN	2310046G15RIK.	GN	DJ223E3.1.
OS	Mus musculus (Mouse).	OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	OC	Eukaryota; Metazoa; Chordata; Craniata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=10090;	[1]	NCBI_TaxID=9006;	[1]
RN	RP SEQUENCE FROM N.A.	RN	SEQUENCE FROM N.A.
RC	STRAIN=C57BL/6J; TISSUE=TONGUE;	RC	SEQUENCE FROM N.A.
RDLINE	MEDLINE=21085660; PubMed=11217851;	RDLINE	SEQUENCE FROM N.A.
DT	01-JUN-2001 (TREMBLrel. 17, Created)	DT	01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT	(TREMBLrel. 17, Last sequence update)	DT	(TREMBLrel. 17, Last sequence update)
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)	DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE	2310046G15RIK PROTEIN.	DE	DJ223E3.1 (PUTATIVE SECRETED PROTEIN ZSIG13).
GN		GN	DJ223E3.1.
RA	Saito T., Okazaki Y., Gojobori T., Bon H., Kasai T., Saito R., Kondo S., Yamanaoka T., Matsuda H.A., Asiburner M., Batalov S., Casavant T., Kiyosawa H., Konno H., Adachi J., Fukuda S., Aizawa K., Iwasa M., Nishi K., Kadota K., Matsuda F., Lewis S., Matsuura T., Gissi C., King B., Kochiwa H., Schriml L.M., Staubli F., Suzuki R., Nikaido I., Pesole G., Quackenbush J., Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G., Blake J., Bojelli D., Bojelli N., Carninci P., de Bonaldo M.F., Brownstein M.J., Bult C., Fletcher C., Fujita M., Garibaldi M., Gustincich S., Hill D., Hoffmann M., Hume D.A., Kamiya M., Lee N.H., Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P., Durm M.; Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.	RA	RA Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
RA	DR EMBL; AL121939; CAC35071.1; -	RA	DR EMBL; AL121939; CAC35071.1; -
RA	SEQUENCE 413 AA; 47098 MW; 818D9C951BD2D6C1 CRC64;	RA	SEQUENCE 413 AA; 47098 MW; 818D9C951BD2D6C1 CRC64;

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

1 protein - protein search, using sw model

Run on: December 2, 2001, 16:16:53 ; Search time 64.47 Seconds
(without alignments)
450.391 Million cell updates/sec

Title: US-09-072-384-15
2131
Sequence: 1 MAGIPGLLFFLLCAGQ.....IKGNYLDREGDTVFPQGSN 392

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing First 45 summaries

Database : A_Geneseq_1101.*

1: /SIDS2/gcdata/geneseq/geneseqp/AA1980.DAT: *
2: /SIDS2/gcdata/geneseq/geneseqp/AA1981.DAT: *
3: /SIDS2/gcdata/geneseq/geneseqp/AA1982.DAT: *
4: /SIDS2/gcdata/geneseq/geneseqp/AA1983.DAT: *
5: /SIDS2/gcdata/geneseq/geneseqp/AA1984.DAT: *
6: /SIDS2/gcdata/geneseq/geneseqp/AA1985.DAT: *
7: /SIDS2/gcdata/geneseq/geneseqp/AA1986.DAT: *
8: /SIDS2/gcdata/geneseq/geneseqp/AA1987.DAT: *
9: /SIDS2/gcdata/geneseq/geneseqp/AA1988.DAT: *
10: /SIDS2/gcdata/geneseq/geneseqp/AA1989.DAT: *
11: /SIDS2/gcdata/geneseq/geneseqp/AA1990.DAT: *
12: /SIDS2/gcdata/geneseq/geneseqp/AA1991.DAT: *
13: /SIDS2/gcdata/geneseq/geneseqp/AA1992.DAT: *
14: /SIDS2/gcdata/geneseq/geneseqp/AA1993.DAT: *
15: /SIDS2/gcdata/geneseq/geneseqp/AA1994.DAT: *
16: /SIDS2/gcdata/geneseq/geneseqp/AA1995.DAT: *
17: /SIDS2/gcdata/geneseq/geneseqp/AA1996.DAT: *
18: /SIDS2/gcdata/geneseq/geneseqp/AA1997.DAT: *
19: /SIDS2/gcdata/geneseq/geneseqp/AA1998.DAT: *
20: /SIDS2/gcdata/geneseq/geneseqp/AA1999.DAT: *
21: /SIDS2/gcdata/geneseq/geneseqp/AA2000.DAT: *
22: /SIDS2/gcdata/geneseq/geneseqp/AA2001.DAT: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Match Length	DB ID	Description
1	2131	100.0	392	22	AAB48973 Human Zsig13 variant
2	2085	97.8	392	22	AAB48972 Human Zsig13 variant
3	2080	97.6	383	20	AAV08657 Human transmembrane protein encoded by W0992704 Seq ID
4	2080	97.6	383	20	AAV13390 Amino acid sequence
5	2080	97.6	383	21	AAB25592 Protein encoded by gene AB25618
6	2080	97.6	383	21	AAB25592 Human TANGO 186 protein
7	2080	97.6	383	21	AAY88277 Human signal peptide
8	2080	97.6	383	21	AAY87270 A bone marrow secreted protein
9	2080	97.6	383	21	AAY53627 Human PRO307 protein
10	2080	97.6	383	22	AAB880258
11	2080	97.6	383	22	AAB880258

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

תְּלִימָדָה

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	2131	100.0	392	22	AAB48973	Human Zsig13 vari
2	2035	97.8	392	22	AAB48972	Human Zsig13 vari
3	2080	97.6	383	20	AAY08557	Human Transmembr
4	2080	97.6	383	20	AAY08650	W0997094 Seq ID
5	2080	97.6	383	20	AAY13390	Amino acid sequen
6	2080	97.6	383	21	AAB25592	Protein encoded b
7	2080	97.6	383	21	AAB25618	rotein encoded b
8	2080	97.6	383	21	AAY88277	Human TANGO_186_P
9	2080	97.6	383	21	AAY88270	Human signal peptide
10	2080	97.6	383	21	AAY53627	A bone marrow pro
11	2080	97.6	383	22	AAB80258	Human PRO307 prob

PT New isolated serine protease (designated Zsig13), useful in industrial processes to degrade unwanted proteins or alter the characteristics of protein-containing composition, as well as in industrial applications (e.g. brewing)

XX

PS Claim 1; Column 35-38; 26pp; English.

CC The invention relates to human Zsig13 proteins (AAB48972-B48974), and CC to DNA encoding them (AAC91782-C91784). The invention also relates to expression vectors and host cells comprising a human Zsig13 DNA, and the recombinant production of a human Zsig13 protein or its precursor.

CC Zsig13 is a serine protease, and has significant homology to Bacillus licheniformis glutamyl endopeptidase, human clotting factor X, human elastase, rat mast cell protease, Streptomyces griseus trypsin, bovine trypsinogen, and Hypoderma lineatum collagenase. The gene encoding human Zsig13 is located on chromosome 11q22.1. Zsig13 is useful in industrial processes to degrade unwanted proteins or alter the characteristics of protein-containing compositions. It may also be used in industrial applications in which proteases are utilised including food processing, brewing and alcohol production, and as a component of a laundry detergent. The present sequence represents a human Zsig13 variant.

XX Sequence 392 AA;

Query Match 100.0%; Score 2131; DB 22; Length 392;
Best Local Similarity 100.0%; Pred. No. 3.8e-152; Gaps 0;
Matches 392; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MAGIPGLFLFLFLCAVGQVSPYSSADPKWPMPAYLPVVLQESTNLAKEPAEAKLE 60
Db 1 magipglflflflcavgqvpssyssadpkwpmpaylpvvlpqstnlakdpfgeakle 60

Qy 61 VSSSCGPQCHKTPPLPYYEAQYLSTETLTLANGSRSETQGYIILSSSGDQAQRHDSG 120
Db 61 vssscgpachktpplpypyeaqylstetlyangsrsetqgyiilsssgdqaqrhdsg 120

Qy 121 SGSSRRKRQIYGDYDSRSFISFGKDFLLNYPPSTSVKLSTGCNGTVAEKHVLAHCITHDG 180
Db 121 sgssrrkrqiyjydsrsifsgdfllnypfsksvk1stgcgtlvaeakhvitaahchdg 180

Qy 181 KTVKGTOKLRLRGFLKFKFDGRGRANDSTAMPMEONKFWTRVKRTHVPGWIKGNND 240
Db 181 ktyvkgqkqkrlrgflkfkfdgrgrandstsampmekfqwirvkrthvpgwikgnnd 240

Qy 241 IGMDDYALLERKPHKRKFMTIGVSPAKQDGGRIHFSGDNDPFGNLVYRECDVKE 300
Db 241 igmddyallekkphkrkfmtigvspakqdpgrifhsgvndrpgnlvlyrcdvke 300

Qy 301 TYDLYXOCDAQPGASGSVYRMWKQQQWERKLTIGIFSHQNYDMNGSFQDFNVAVR 360
Db 301 tydlyxocdaqpgasgsqvrmwkqqkwerkilgifsqhwdmngsfqdfnavr 360

Qy 361 ITPLKYAQICYWIKGNYLDCREGDTYFPPGSN 392
Db 361 itplkyaqicywikgnyldcrgdtvfppgsn 392

RESULT 2
ID AAB48972
XX AAB48972 standard; Protein; 392 AA.
AC AAB48972;
XX DT 27-MAR-2001 (first entry)
DE Human Zsig13 variant #1, SEQ ID NO:2
XX

KW Human Zsig13; serine protease; chromosome 11q22.1; elastase homologue; glutamyl endopeptidase; factor X homologue; trypsin homologue; trypsinogen homologue; mast cell protease homologue; collagenase homologue; protein degradation; food processing; brewing; alcohol production; laundry detergent component.

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2001, 16:15:34 ; Search time 35.78 Seconds

(without alignments)
240.882 million cell updates/sec

Title: US-09-072-384-18

Perfect score: 2080

Sequence: 1 MAGIPGLLFLLLFLCAVGQ.....LKYAQICYWIKGNYLDREG 383

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA:
 1: /cgn2_6/ptodata/2/1aa/5A_COMBO.pep:/*
 2: /cgn2_6/ptodata/2/1aa/5B_COMBO.pep:/*
 3: /cgn2_6/ptodata/2/1aa/6A_COMBO.pep:/*
 4: /cgn2_6/ptodata/2/1aa/6B_COMBO.pep:/*
 5: /cgn2_6/ptodata/2/1aa/PCRTUS_COMBO.pep:/*
 6: /cgn2_6/ptodata/2/1aa/backfile1.pep:/*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2080	100.0	383	4 US-09-072-384-18	Sequence 18, Appl
2	2080	100.0	392	4 US-09-072-384-15	Sequence 15, Appl
3	2044	98.3	392	4 US-09-072-384-2	Sequence 1, Appl
4	177	8.5	222	1 US-08-090-048-1	Sequence 1, Appl
5	177	8.5	222	2 US-08-292-550-1	Sequence 1, Appl
6	177	8.5	222	2 US-17-922-661A-1	Sequence 1, Appl
7	122	5.9	256	3 US-08-906-769-89	Sequence 89, Appl
8	122	5.9	256	3 US-08-906-616-89	Sequence 89, Appl
9	122	5.9	256	4 US-08-817-795-89	Sequence 89, Appl
10	122	5.9	256	4 US-08-639-075A-89	Sequence 89, Appl
11	122	5.9	256	4 US-09-012-431-89	Sequence 89, Appl
12	122	5.9	256	4 US-09-012-215-32	Sequence 89, Appl
13	122	5.9	256	4 US-09-012-692-89	Sequence 89, Appl
14	122	5.9	256	4 US-08-906-613-89	Sequence 89, Appl
15	122	5.9	256	5 PCT-US95-14412A-89	Sequence 2, Appl
16	117	5.6	437	1 US-08-481-037-2	Sequence 59, Appl
17	115	5.6	241	4 US-08-908-483-59	Sequence 44, Appl
18	114	5.5	228	4 US-08-914-483-44	Patient No. 5223425
19	114	5.5	253	6 5223425-8	Patient No. 5223425
20	113	5.4	238	6 5223425-5	Patient No. 5223425
21	113	5.4	250	6 5223425-4	Patient No. 5223425
22	112.5	5.4	223	1 US-08-298-091-13	Sequence 13, Appl
23	112.5	5.4	223	1 US-08-433-859-13	Sequence 13, Appl
24	112.5	5.4	223	2 US-08-471-167-13	Sequence 13, Appl
25	112.5	5.4	223	2 US-08-492-816-13	Sequence 13, Appl
26	112.5	5.4	223	2 US-08-296-149-13	Sequence 13, Appl

ALIGNMENTS

RESULT 1
US-09-072-384-18

; Sequence 18, Application US/09072384

; Patent No. 615420

; GENERAL INFORMATION:

; APPLICANT: Sheppard, Paul O.

; TITLE OF INVENTION: SERINE PROTEASE POLYPEPTIDES

; NUMBER OF SEQUENCES: 18

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: ZymoGenetics, Inc.

; STREET: 1201 Eastlake Avenue East

; CITY: Seattle

; STATE: WA

; COUNTRY: USA

; ZIP: 98102

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSEQ for Windows Version 2.0

; CURRENT APPLICATION DATA:

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION NUMBER:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Parker, Gary E.

; REGISTRATION NUMBER: 31,648

; REFERENCE/DOCKET NUMBER: 97-16C1

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 206-442-6673

; TELEFAX: 206-442-6678

; INFORMATION FOR SEQ ID NO: 18:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 383 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; FRAGMENT TYPE: internal

; NAME/KEY: Signal Sequence

; LOCATION: 1...19

; OTHER INFORMATION:

US-09-072-384-18

Query Match 100.0%; Score 2080; DB 4; Length 383;
 Best Local Similarity 100.0%; Pred. No. 5e-223; Gaps 0;
 Matches 383; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MAGIPGLLFLFFLICAVQSPYSPAPWKPTWPAYRLPVVLPOSTLNIAKPDFGAEAKLE 60
 Db 1 MAGIPGLLFLFFLICAVQSPYSPAPWKPTWPAYRLPVVLPOSTLNIAKPDFGAEAKLE 60

Qy 61 VSSSGPQCHKGTPILPTYEAKQQLSYETLYANSRTERQVGYIYLSSSGDGAQHRDGS 120
 Db 61 VSSSGPQCHKGTPILPTYEAKQQLSYETLYANSRTERQVGYIYLSSSGDGAQHRDGS 120

Qy 121 SGSSRKQRQIYGYDSRSFESLGKDFLLNPFTSYKLSPGTGTLVAEKHVLTAHCIDHG 180
 Db 121 SGSSRKQRQIYGYDSRSFESLGKDFLLNPFTSYKLSPGTGTLVAEKHVLTAHCIDHG 180

Qy 181 KTYVGTQKLRGVGFPLPKFKDGGGRANDSTSAMPQMKFWIRVKRTHVPKGWIKGNAND 24.0
 Db 181 KTYVGTQKLRGVGFPLPKFKDGGGRANDSTSAMPQMKFWIRVKRTHVPKGWIKGNAND 24.0

Qy 241 GMDDYALLELKPHKRKEMKIGCVSPPAKQPLGPGRINFSGYDRPGNLYRFDYKDE 300
 Db 241 GMDDYALLELKPHKRKEMKIGCVSPPAKQPLGPGRINFSGYDRPGNLYRFDYKDE 300

Qy 301 TYDLQQCDAQPAGASGSGVYVRMWWKROQQKWERKERIGIFSGHQWDMMNGSPQDENYAVR 360
 Db 301 TYDLQQCDAQPAGASGSGVYVRMWWKROQQKWERKERIGIFSGHQWDMMNGSPQDENYAVR 360

Qy 361 ITPLKYAQICYWIGNYLDREG 383
 Db 361 ITPLKYAQICYWIGNYLDREG 383

RESULT 2
 US-09-072-384-15
 Sequence 15, Application US/09072384
 Patent No. 6153420

GENERAL INFORMATION:
 APPLICANT: Sheppard, Paul O.
 TITLE OF INVENTION: SERINE PROTEASE POLYPEPTIDES
 TITLE OF INVENTION: AND MATERIALS AND METHODS FOR MAKING THEM
 NUMBER OF SEQUENCES: 18
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: ZymoGenetics, Inc.
 STREET: 1201 Eastlake Avenue East
 CITY: Seattle
 STATE: WA
 COUNTRY: USA
 ZIP: 98102

COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM compatible
 OPERATING SYSTEM: DOS
 SOFTWARE: FASTSEQ for Windows Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/072,384
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER:
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: PERET, GARY E
 REGISTRATION NUMBER: 31,648
 REFERENCE/DOCKET NUMBER: 97-16C1
 TELEPHONE: 206-442-6673
 TELEX: 206-442-6678
 INFORMATION FOR SEQ ID NO: 15:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 392 amino acids

TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 FRAGMENT TYPE: internal
 FEATURE:
 NAME/KEY: Signal Sequence
 LOCATION: 1..19
 OTHER INFORMATION:
 US-09-072-384-15

Query Match 100.0%; Score 2080; DB 4; Length 392;
 Best Local Similarity 100.0%; Pred. No. 5.e-223; Gaps 0;
 Matches 383; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MAGIPGLLFLFFLICAVQSPYSPAPWKPTWPAYRLPVVLPOSTLNIAKPDFGAEAKLE 60
 Db 1 MAGIPGLLFLFFLICAVQSPYSPAPWKPTWPAYRLPVVLPOSTLNIAKPDFGAEAKLE 60

Qy 61 VSSSGPQCHKGTPILPTYEAKQQLSYETLYANSRTERQVGYIYLSSSGDGAQHRDGS 120

Query Match 100.0%; Score 2080; DB 4; Length 392;
 Best Local Similarity 100.0%; Pred. No. 5.e-223; Gaps 0;
 Matches 383; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MAGIPGLLFLFFLICAVQSPYSPAPWKPTWPAYRLPVVLPOSTLNIAKPDFGAEAKLE 60
 Db 1 MAGIPGLLFLFFLICAVQSPYSPAPWKPTWPAYRLPVVLPOSTLNIAKPDFGAEAKLE 60

Qy 61 VSSSGPQCHKGTPILPTYEAKQQLSYETLYANSRTERQVGYIYLSSSGDGAQHRDGS 120

Db 61 VSSSGPQCHKGTPILPTYEAKQQLSYETLYANSRTERQVGYIYLSSSGDGAQHRDGS 120

Qy 121 SGSSRKQRQIYGYDSRSFESLGKDFLLNPFTSYKLSPGTGTLVAEKHVLTAHCIDHG 180
 Db 121 SGSSRKQRQIYGYDSRSFESLGKDFLLNPFTSYKLSPGTGTLVAEKHVLTAHCIDHG 180

Qy 181 KTYVGTQKLRGVGFPLPKFKDGGGRANDSTSAMPQMKFWIRVKRTHVPKGWIKGNAND 24.0
 Db 181 KTYVGTQKLRGVGFPLPKFKDGGGRANDSTSAMPQMKFWIRVKRTHVPKGWIKGNAND 24.0

Qy 241 GMDDYALLELKPHKRKEMKIGCVSPPAKQPLGPGRINFSGYDRPGNLYRFDYKDE 300
 Db 241 GMDDYALLELKPHKRKEMKIGCVSPPAKQPLGPGRINFSGYDRPGNLYRFDYKDE 300

Qy 301 TYDLQQCDAQPAGASGSGVYVRMWWKROQQKWERKERIGIFSGHQWDMMNGSPQDENYAVR 360
 Db 301 TYDLQQCDAQPAGASGSGVYVRMWWKROQQKWERKERIGIFSGHQWDMMNGSPQDENYAVR 360

Qy 361 ITPLKYAQICYWIGNYLDREG 383
 Db 361 ITPLKYAQICYWIGNYLDREG 383

RESULT 3
 US-09-072-384-2
 Sequence 2, Application US/09072384
 Patent No. 6153420

GENERAL INFORMATION:
 APPLICANT: Sheppard, Paul O.
 TITLE OF INVENTION: SERINE PROTEASE POLYPEPTIDES
 NUMBER OF SEQUENCES: 18
 CORRESPONDENCE ADDRESS:
 ADDRESS: ZymoGenetics, Inc.
 STREET: 1201 Eastlake Avenue East
 CITY: Seattle
 STATE: WA
 COUNTRY: USA
 ZIP: 98102

COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM compatible
 OPERATING SYSTEM: DOS
 SOFTWARE: FASTSEQ for Windows Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/072,384
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER:
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: PERET, GARY E
 REGISTRATION NUMBER: 31,648
 REFERENCE/DOCKET NUMBER: 97-16C1
 TELEPHONE: 206-442-6673
 TELEX: 206-442-6678
 INFORMATION FOR SEQ ID NO: 15:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 392 amino acids

PT Novel proteins containing transmembrane domains, useful as
 PT anti-inflammatory, immune stimulators/suppressors and tissue
 PR growth compounds
 XX
 PS Claim 1; Page 68-69; 89pp; English.
 XX
 CC This invention describes novel human transmembrane containing proteins
 CC and their encoding nucleic acids. Although no specific use is given for
 CC the proteins, they may have a range of activities selected from
 CC nutritional uses, cytokine and cell differentiation, immune
 CC stimulation/suppression, haemopoiesis regulatory, tissue growth,
 CC activin/inhibin, chemostatic/chemokinetic, haemostatic/thrombolytic,
 CC receptor/ligand, tumour inhibitor, anti-inflammatory and other undefined
 CC activities. The cDNAs can be utilized as probes for gene diagnosis and
 CC expression of proteins. The transformed cells can be used for large
 CC scale detection of the corresponding ligands and for screening of novel
 CC low-molecular pharmaceuticals.
 XX
 Sequence 383 AA;

Query Match 100.0%; Score 2080; DB 20; Length 383;
 Best Local Similarity 100.0%; Pred. No. 1e-146;
 Matches 383; Conservative 0; Mismatches 0; Gaps 0;
 Indels 0;
 1 MAGIPGLLFLLLFLCAGVQVSPPYSAPAKPPTNPAYRLPVVLPOSTLNIAKPDFGAEAKLE 60
 Db 1 magipglflflflcavgvqvsppysapakpptaipvvlpipstlnakpdfgaaekle 60
 Qy 61 VSSCGPPOCHKGIPPLPYBEAKOYLSTETLYANGSRTEPQGYIILSSSGDQAQHRSGS 120
 Db 61 vssscgpqchkgipplpybeakoylstetlyangsrtepqtyiilsssgdqaqhrgs 120
 Qy 121 SGKSRRKRQIYGYDSRSTIEFGKDFLLNYPFSTSIVKLSTGTCTGTLVAEKHVTAAHCITHDG 180
 Db 121 sgssrrkrqiygydsrstiefgkdfllnypfstsivklstgtctgtlvaevhvtahchhdg 180
 Qy 181 KTVVKGTQKLRVGLPKPKFKDGGRGANDSTSAMPQMKFQWIVKTRHVKRHPWKVKGNAWD 240
 Db 181 ktvvkgtqkrlrvglpkfkfdggrrganstsampqmkfqwivkrtrhvkrhpwkvgna 240
 Qy 241 IGMDDYDALLEKKPHKAKEMKIGVSPPAKQPLGGRTHFSGGTDNDRPGNLVYRCDVYDE 300
 Db 241 igmddydalekkphkakemkigvspapkqplggrthfsggtndrpgnlvyrcdvde 300
 Qy 301 TYDILYQQCDAQPGASGSGVYMWKROQQKWRKIGIFSGSHOWDMNGSSQDFNVAVR 360
 Db 301 tydillyqqcdaqpgasgsgvymwkkroqqkwrkigifsgshowdmngssqdfnvavr 360
 Qy 361 ITPLKYAOTCYNKGNYLDCREG 383
 Db 361 itplkyadicywkgnyldcreg 383
 RESULT 2
 ID AAY08660 Standard; Protein; 383 AA.
 XX
 AAY08660;
 XX
 DT 09-AUG-1999 (first entry)
 XX
 DE W09927094 Seq ID 12.
 XX
 Transmembrane domain; human; nutrition; cytokine; cell differentiation;
 KW immune stimulation; immune suppression; haemopoiesis; activin;
 KW regulatory tissue growth; inhibin; chemostatic; chemokinetic;
 KW haemostatic; thrombolytic; tumour inhibitor; anti-inflammatory;
 KW gene therapy; screening.
 XX
 Homo sapiens.
 OS

PN W09927094-A2.
 XX
 PD 03-JUN-1999.
 XX
 PF 20-NOV-1998; 98WO-JP05238.
 XX
 PR 25-NOV-1997; 97JP-0323129.
 XX
 PA (PROT-) PROTEGENE INC.
 PA (SAGA) SAGAMI CHEM RES CENT.
 XX
 Kato S, Kimura T, Sekine S;
 PI
 DR WPI: 1999-357835/30.
 XX
 Novel proteins containing transmembrane domains, useful as
 PT anti-inflammatories, immune stimulators/suppressors and tissue
 PT growth compounds
 PT Disclosure; Page 87-89; 89pp; English.
 XX
 This invention describes novel human transmembrane containing proteins
 CC and their encoding nucleic acids. Although no specific use is given for
 CC the proteins, they may have a range of activities selected from
 CC nutritional uses, cytokine and cell differentiation, immune
 CC stimulation/suppression, haemopoiesis regulatory, tissue growth,
 CC activin/inhibin, chemostatic/chemokinetic, haemostatic/thrombolytic,
 CC receptor/ligand, tumour inhibitor, anti-inflammatory and other undefined
 CC activities. The cDNAs can be utilized as probes for gene diagnosis and
 CC as gene sources for gene therapy. The cDNAs can also be used for large
 CC scale expression of proteins. The transformed cells can be used for
 CC detection of the corresponding ligands and for screening of novel
 CC low-molecular pharmaceuticals.
 XX
 Sequence 383 AA;

Query Match 100.0%; Score 2080; DB 20;
 Best Local Similarity 100.0%; Pred. No. 1e-146;
 Matches 383; Conservative 0; Mismatches 0; Gaps 0;
 Indels 0;
 Qy 1 MAGIPGLLFLLLFLCAGVQVSPPYSAPAKPPTNPAYRLPVVLPOSTLNIAKPDFGAEAKLE 60
 Db 1 magipglflflflcavgvqvsppysapakpptaipvvlpipstlnakpdfgaaekle 60
 Qy 61 VSSCGPPOCHKGIPPLPYBEAKOYLSTETLYANGSRTEPQGYIILSSSGDQAQHRSGS 120
 Db 61 vssscgpqchkgipplpybeakoylstetlyangsrtepqtyiilsssgdqaqhrgs 120
 Qy 121 SGKSRRKRQIYGYDSRSTIEFGKDFLLNYPFSTSIVKLSTGTCTGTLVAEKHVTAAHCITHDG 180
 Db 121 sgssrrkrqiygydsrstiefgkdfllnypfstsivklstgtctgtlvaevhvtahchhdg 180
 Qy 181 KTVVKGTQKLRVGLPKPKFKDGGRGANDSTSAMPQMKFQWIVKTRHVKRHPWKVKGNAWD 240
 Db 181 ktvvkgtqkrlrvglpkfkfdggrrganstsampqmkfqwivkrtrhvkrhpwkvgna 240
 Qy 241 IGMDDYDALLEKKPHKAKEMKIGVSPPAKQPLGGRTHFSGGTDNDRPGNLVYRCDVYDE 300
 Db 241 igmddydalekkphkakemkigvspapkqplggrthfsggtndrpgnlvyrcdvde 300
 Qy 301 TYDILYQQCDAQPGASGSGVYMWKROQQKWRKIGIFSGSHOWDMNGSSQDFNVAVR 360
 Db 301 tydillyqqcdaqpgasgsgvymwkkroqqkwrkigifsgshowdmngssqdfnvavr 360
 Qy 361 ITPLKYAOTCYNKGNYLDCREG 383
 Db 361 itplkyadicywkgnyldcreg 383
 RESULT 2
 ID AAY13390 Standard; Protein; 383 AA.
 XX
 AAY13390 standard; Protein; 383 AA.

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2001, 16:25:50 ; Search time 84.84 Seconds
(without alignments)
660,329 Million cell updates/sec

Title: US-09-072-384-18
Perfect score: 2080
Sequence: 1 MAGIPGFLFLFLCAVGO.....LKYAQICYWIKGNYLDREG 383

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 0%

Listing first 45 summaries

Database : SPTREMBL_17:

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rabbit:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2080	100.0	383	4 O95084	O95084 homo sapien
2	1884.5	90.6	382	1 O9D6X6	O9d6x6 mus musculu
3	1042	50.1	413	4 Q9BQP6	Q9bqp6 homo sapien
4	129.5	6.4	799	11 Q9DBI0	Q9dbi0 mus musculu
5	129.5	6.2	303	2 Q9EXR9	Q9exr9 bacillus in
6	129	6.2	469	6 Q9GMD9	Q9gmd9 ornithorhyn
7	127.5	6.1	1322	5 Q9NATO	Q9nato anopheles g
8	126.5	6.1	678	11 Q9JTS8	Q9jts8 ratus norv
9	124.5	6.0	1322	5 Q9NJS5	Q9njss anopheles g
10	124	6.0	266	6 O46644	O46644 macaca fasc
11	123.5	5.9	339	11 Q9QX91	Q9qx91 rattus norv
12	123.5	5.9	366	11 Q9QX85	Q9qx85 rattus norv
13	123.5	5.9	541	11 Q9QX90	Q9qx90 rattus norv
14	123.5	5.9	623	11 Q9JJP3	Q9jjp3 rattus norv
15	123.5	5.9	643	11 Q9QX84	Q9qx84 rattus norv
16	122	5.9	259	5 Q9XY61	Q9xy61 ctenocephal
17	122	5.9	449	5 Q9VDU8	Q9vdub drosophila
18	120.5	5.8	482	11 Q63207	Q63207 rattus norv
19	118	5.7	1376	5 Q9VQR8	Q9vqr8 drosophila

ALIGNMENTS

20	117	5.6	258	5 Q9W5U8	Q9w5u8 drosophila
21	117	5.6	522	5 Q9D936	Q9d936 mus musculu
					Q14243 homo sapien
					Q9un11 homo sapien
					Q9va87 drosophila
					Q9z338 mus musculu
					Q9q122 brachydanio
					Q9v516 drosophila
					Q9vdv1 drosophila
					Q9bz10 homo sapien
					Q9w9q1 paralichthys sex
					Q96991 manduca sexta
					Q54740 mus musculu
					Q10922 caenorhabditis
					Q69973 streptomyces
					Q170866 anophelles s
					Q9w633 cyprinus carpio
					Q24019 drosophila
					Q9z1h1 mus musculu
					Q9ay44 chaetoceros
					Q88947 mus musculu
					Q99132 mus musculu
					Q9vz88 drosophila
					Q9vfz6 drosophila
					Q9yic6 cyprinus carpio

28	113	5.4	573	5 Q9V516	Q9v516 drosophila
29	112.5	5.4	603	5 Q9VDV1	Q9vdv1 drosophila
30	112.5	5.4	686	4 Q9BZHO	Q9bz10 homo sapien
31	112	5.4	249	13 Q9W7Q1	Q9w9q1 paralichthys sex
32	110.5	5.3	279	5 Q96991	Q96991 manduca sexta
33	110.5	5.3	481	11 Q54740	Q54740 mus musculu
34	110	5.3	1449	5 Q10922	Q10922 caenorhabditis
35	109.5	5.3	405	2 Q69973	Q69973 streptomyces
36	108	5.2	274	5 Q170866	Q170866 anophelles s
37	108	5.2	750	13 Q9W633	Q9w633 cyprinus carpio
38	108	5.2	1047	5 Q24019	Q24019 drosophila
39	107.5	5.2	236	11 Q9Z1H1	Q9z1h1 mus musculu
40	107.5	5.2	427	10 Q9AYRA4	Q9ay44 chaetoceros
41	107.5	5.2	481	11 Q88947	Q88947 mus musculu
42	107.5	5.2	481	11 Q99132	Q99132 mus musculu
43	107.5	5.2	611	5 Q9VZS8	Q9vzs8 drosophila
44	106	5.1	460	5 Q9VFZ6	Q9vfz6 drosophila
45	106	5.1	762	13 Q9YIC6	Q9yic6 cyprinus carpio

RESULTS	1				
095084	ID				
	AC				
	DT	01-MAY-1999	(TREMBLrel. 1.0, Created)		
	DT	01-MAY-1999	(TREMBLrel. 1.0, Last sequence update)		
	DT	01-JUN-2001	(TREMBLrel. 1.7, Last annotation update)		
	DB		SERINE PROTEASE (HYPOTHETICAL 43.0 KDA PROTEIN) (PROTEASE, SERINE, ZSIG13 OR DKFZP586B0719).		
	DE				
	GN				
	OS		Homo sapiens (Human), Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
	OC				
	RT		"A novel serine protease from human umbilical vein endothelial cells.";		
	RT		NCBI TAXID=9606;		
	RN		[1]		
	RP		SEQUENCE FROM N.A.		
	RC		TISSUE=UMBILICAL VEIN;		
	RA		Sheppard P., Blumberg H., Jelinek L., Foster D., O'Hara P.; Li X., Tedder T.F.;		
	RT		"A novel serine protease from human umbilical vein endothelial cells.";		
	RT		Submitted (JUL-1997) to the EMBL/GenBank/DDBJ databases.		
	RN		[1]		
	RP		SEQUENCE FROM N.A.		
	RC		TISSUE=UTERUS;		
	RA		Wambutt R., Reubner D., Newes H.W., Gassenhuber J., Wiemann S.;		
	RT		Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.		
	RN		[3]		
	RP		SEQUENCE FROM N.A.		
	RC		TISSUE=CERVIX CARCINOMA;		
	RA		Strausberg R.,		
	RL		Submitted (DEC-2000) to the EMBL/GenBank/DDBJ databases.		
	CC	-1	SIMILARITY: TO SERINE PROTEASES, TRIPSIN FAMILY.		
	CC	-1	SIMILARITY: TO CHYMOTRYPSIN SERINE PROTEASE FAMILY (S1).		
	EMBL	AFO15287; AAD0153.1;			
	DR	AF193611; AAF07186.1;			
	DR	EML; AL166914; CAB6648.1;			
	DR	BC001278; AAH01278.1;			
	DR	MEROPS; S01.309;			

DR	InterPro; IPR001314; Chymotrypsin.	RA	Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N., Saeki H., Sato K., Schoenbach C., Seya T., Shiba Y., Storch K.-F., Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilmung L., Wyrshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S., Hayashizaki Y.;
DR	PF00089; trypsin; 1.	RA	"Functional annotation of a full-length mouse cDNA collection.";
SMART	PRO00722; CHYMOTRYPSIN.	RT	Nature 409:685-690(2001).
DR	SMART; SM00020; TRY_SPEC; 1.	RL	-1 - SIMILARITY: TO CHYMOTRYPSIN SERINE PROTEASE FAMILY.
PROSITE	PS00134; TRYPSIN_HIS; UNKNOWN_1.	CC	CC -1 - SIMILARITY: TO CHYMOTRYPSIN SERINE PROTEASE FAMILY (S1).
KW	Hydrolase; Proteases; Serine protease.	DR	EMBL; AK009847; BAB26541.1;
SEQUENCE	383 AA; 43001 MW; 4666C11ABFD5E8F CRC64;	DR	MGDB; MG1:1923703; 3130046G15Rik.
Query Match	100.0%; Score 2080; DB 4; Length 383;	DR	InterPro; IP0001314; Chymotrypsin.
Best Local Similarity	100.0%; Pred. No. 8_2e-14;	DR	InterPro; IP0001254; Trypsin.
Matches	383; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	DR	Pfam; PF00089; trypsin; 1.
Qy	1 MAGIPGLLFLFLFLCAVQVSPYSAPWKPTNPAYRLPVLPVLPVPOSTLNLAKEAAKLE 60	DR	PRINTS; PR00022; CHYMOTRYPSIN.
Db	1 MAGIPGLLFLFLFLCAVQVSPYSAPWKPTNPAYRLPVLPVLPVPOSTLNLAKEAAKLE 60	DR	SM00020; TRY_SPEC; 1.
Qy	61 VSSSCGPQCHKGTPPLTYEEAKOYLSETYLYANGSRTETOVGYIYLSSGDGAQHRDSGS 120	DR	PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
Db	61 VSSSCGPQCHKGTPPLTYEEAKOYLSETYLYANGSRTETOVGYIYLSSGDGAQHRDSGS 120	KW	Hydrolase; Serine protease.
Query	90.6%; Score 1884.5; DB 11; Length 382;	SEQUENCE	382 AA; 43053 MW; 69C9A7080E5B2306 CRC64;
Best Local Similarity	90.6%; Pred. No. 9_2e-166;	Query Match	90.6%; Score 1884.5;
Matches	347; Conservative 15; Mismatches 20; Indels 1; Gaps 1;	Best Local Similarity	90.6%; Pred. No. 9_2e-166;
Qy	1 MAGIPGLLFLFLFLCAVQVSPYSAPWKPTNPAYRLPVLPVLPVPOSTLNLAKEAAKLE 60	Matches	347; Conservative 15; Mismatches 20; Indels 1; Gaps 1;
Db	1 MAGIPGLLFLFLFLCAVQVSPYSAPWKPTNPAYRLPVLPVLPVPOSTLNLAKEAAKLE 60	Qy	1 MAGIPGLLFLFLFLCAVQVSPYSAPWKPTNPAYRLPVLPVLPVPOSTLNLAKEAAKLE 60
Qy	181 KTYVKGTQKLRLQFLKPKFKDGRGRANDSTSAMPQMKFOWLRVKRTHVPKGWIKGNAND 240	Db	1 MAGIPGLLFLFLFLCAVQVSPYSAPWKPTNPAYRLPVLPVLPVPOSTLNLAKEAAKLE 59
Db	181 KTYVKGTQKLRLQFLKPKFKDGRGRANDSTSAMPQMKFOWLRVKRTHVPKGWIKGNAND 240	Qy	61 VSSSCGPQCHKGTPPLTYEEAKOYLSETYLYANGSRTETOVGYIYLSSGDGAQHRDSGS 120
Qy	241 IGMDYDALLELKPKHKKFMIGVSPAKQLPGRTRHSGYTDNDREGNLYRFCDVYKDE 300	Db	60 VSSSCGPQCHKGTPPLTYEEAKOYLSETYLYANGSRTETOVGYIYLSSGDGAQHRDSGS 119
Db	241 IGMDYDALLELKPKHKKFMIGVSPAKQLPGRTRHSGYTDNDREGNLYRFCDVYKDE 300	Qy	121 SGNSRKRTQYGDTSRSTFGKDFLINYPSVSVKLSTGCTLVAEKHVTAACHTHDG 180
Qy	301 TYDLYQQCDAQPGASGSVYRMWKRQQQKWERKLIGIFSGHWDNGSSQDENAVVR 360	Db	120 TGRSRRRKQTQYGDGRSIFGKDFLINYPSVSVKLSTGCTLVAEKHVTAACHTHDG 179
Db	301 TYDLYQQCDAQPGASGSVYRMWKRQQQKWERKLIGIFSGHWDNGSSQDENAVVR 360	Qy	181 KTYVKGTQKLRLQFLKPKFKDGRGRANDSTSAMPQMKFOWLRVKRTHVPKGWIKGNAND 240
Qy	361 ITPLKYAQICWYKGNYLDREG 383	Db	180 KHYVKGPKQKLRLQFLKPKFKDGRGRANDSTSAMPQMKFOWLRVKRTHVPKGWIKGNAND 239
Db	361 ITPLKYAQICWYKGNYLDREG 383	Qy	121 IGMDYDALLELKPKHKKFMIGVSPAKQLPGRTRHSGYTDNDREGNLYRFCDVYKDE 300
RESULT	2	Db	120 TGRSRRRKQTQYGDGRSIFGKDFLINYPSVSVKLSTGCTLVAEKHVTAACHTHDG 299
Q9D6X6	PRELIMINARY; PRT; 382 AA.	Qy	241 IGMDYDALLELKPKHKKFMIGVSPAKQLPGRTRHSGYTDNDREGNLYRFCDVYKDE 300
ID	AC Q9D6X6;	Db	300 TYDLYQQCDAQPGASGSVYRMWKRQQQKWERKLIGIFSGHWDNGSSQDENAVVR 359
AC	PRELIMINARY; PRT; 382 AA.	Qy	361 ITPLKYAQICWYKGNYLDREG 383
DT	01-JUN-2001 (TREMBLrel. 17, Created)	Db	360 ITPLKYAQICWYKGNYLDREG 382
DT	(TREMBLrel. 17, Last sequence update)	RN	SEQUENCE FROM N.A.
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)	ID	O9BPQ6
DE	DE	AC	O9BPQ6; PRELIMINARY; PRT; 382 AA.
DE	DE	DT	01-JUN-2001 (TREMBLrel. 17, Created)
GN	GN	DT	01-JUN-2001 (TREMBLrel. 17, Last sequence update)
OS	OS	DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	DE	D0223E3.1 (PUTATIVE SECRETED PROTEIN ZSIG13).
Mammalia; Eutheria; Rodentia; Muridae; Murinae; Mus.	Mammalia; Eutheria; Rodentia; Muridae; Murinae; Mus.	GN	D0223E3.1.
OX	OX	GN	Homo sapiens (Human).
NCBI_TaxID=10000;	NCBI_TaxID=10000;	OC	OC
RN	RN	OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
SEQUENCE FROM N.A.	SEQUENCE FROM N.A.	OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RC	RC	OC	NCBI_TaxID=9606;
STRAIN=C57BL/6J; TISSUE=TONGUE;	STRAIN=C57BL/6J; TISSUE=TONGUE;	OX	SEQUENCE FROM N.A.
MEDLINE=21085660; PubMed=11217851;	RA	RA	Dunn M.;
Kawai J., Shinagawa A., Shiba K., Yoshino M., Itoh M., Ishii Y., Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S., Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamamoto I., Saito T., Okazaki Y., Gojobori T., Batalov S., Casavant T., Matsuda H.A., Ashburner M., Bono H., Kasukawa T., Saito R., Kadota K., Gaastrland T., Gissi C., King B., Kochiwa H., Fleischmann W., Lewis S., Matsuo Y., Nikaido I., Pesse L., Schriml L.M., Stabili F., Suzuki R., Tomita M., Wagner L., Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G., Blaize J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F., Brownstein M.J., Bult C., Fletcher C., Fujita M., Garibaldi M., Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H., Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P., RA	RA	RA	
Brownstein M.J., Bult C., Fletcher C., Fujita M., Garibaldi M., Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H., Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P., RA	RA	RA	
Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.	RA	RA	Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
EMBL; AL121939; CAC35071.1;	RA	RA	EMBL; AL121939; CAC35071.1;
SEQUENCE - 413 AA; 41098 MW;	RA	RA	SEQUENCE - 413 AA; 41098 MW; 818D9C951BD2D6C1 CRC64;

Copyright GenCore version 4.5
 (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2001, 16:24:07 ; Search time 32.13 Seconds
 (without alignments)

Perfect score: 2131
 Sequence: 1 MAPIGLLFFLLCAVGQ IKGNYLDCREGDTVFFPGSN 392

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08
 Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39+*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Match	Length	DB ID	Description
1	194	9.1	316	1	GSPB_BACLI	P80057 bacillus li
2	135.5	6.4	313	1	MPR_BACSU	P39790 bacillus su
3	124	5.8	266	1	ELI_BOVIN	Q28153 bos taurus
4	121	5.7	490	1	FA10_RABIT	Q19045 oryctolagus
5	119	5.6	269	1	EL2_PIG	P04419 sus scrofa
6	115.5	5.4	269	1	EL2A_HUMAN	P08217 homo sapien
7	115	5.4	271	1	EL2_MOUSE	P05208 mus musculus
8	115	5.4	376	1	FA10_TROCA	P81428 tropidectis
9	114.5	5.4	238	1	TRY5_AEDEAE	P29787 aedes aegypti
10	113	5.3	266	1	EL1_PIG	P00772 sus scrofa
11	113	5.3	266	1	EL1_RAT	P00773 rattus norvegicus
12	112.5	5.3	259	1	TRYP_STGGR	P07775 streptomyces
13	112.5	5.3	273	1	YGD_ECOLI	P76176 escherichia
14	111.5	5.2	488	1	FA10_HUMAN	P0742 homo sapien
15	111.5	5.2	875	1	NEFR_HUMAN	P56730 homo sapien
16	111	5.2	253	1	CFAD_HUMAN	P00746 homo sapien
17	111	5.2	761	1	NEFR_MOUSE	Q08762 mus musculus
18	110	5.2	269	1	EL2B_HUMAN	P08218 homo sapien
19	110	5.2	492	1	FA10_BOVIN	P0743 bios taurinus
20	108	5.1	583	1	CFAL_HUMAN	P05156 homo sapien
21	108	5.1	786	1	STOB_DRONE	P03319 drosophila
22	107.5	5.0	686	1	MAS2_HUMAN	Q00187 homo sapien
23	107	5.0	269	1	EL2_BOVIN	P29461 bos taurus
24	107	5.0	603	1	CFAL_MOUSE	Q81129 mus musculus
25	105.5	5.0	604	1	CFAL_RAT	Q9uw3 ratius norvegicus
26	105	4.9	274	1	TRY5_ANOGA	P35039 anophelis gambiae
27	104.5	4.9	268	1	TRYP_STGGA	Q54179 streptomyces
28	103.5	4.9	271	1	EL2_RAT	P00774 ratius norvegicus
29	103.5	4.9	416	1	FA10_BOVIN	P00741 bos taurus
30	103	4.8	855	1	ST14_HUMAN	Q9y56 homo sapien
31	102	4.8	855	1	ST14_MOUSE	P56677 mus musculus
32	101.5	4.8	281	1	TRYZ_DRDR	P54630 drosophila
33	98.5	4.6	260	1	COGS_HYPLT	P8897 hypoderma 1

ALIGNMENTS

RESULT	1	GSEPBACLI	STANDARD;	PRT;	316 AA.
ID	GSEPBACLI				
AC	P80057;				
DT	01-MAY-1992 (Rel. 22, Created)				
DT	01-OCT-1993 (Rel. 27, Last sequence update)				
DT	20-AUG-2001 (Rel. 40, Last annotation update)				
DE	GLUTAMYL ENDOPETIDASE PRECURSOR (EC 3.4.21.19) (GLUTAMATE SPECIFIC ENDOPEPTIDASE) (GSE).				
GN	Bacillus licheniformis				
OS	Bacteria; Firmicutes; Clostridium group;				
OC	Bacteria; Firmicutes; Staphylococcus group; Bacillus.				
OX					
RN	[1]				
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.				
RC	STRAIN=ATCC 14580;				
RX	MEDLINE=93054737; PubMed=1429718;				
RA	Kakudo S., Kikuchi N., Kitadokoro K., Fujiwara T., Nakamura E.,				
RA	Okamoto H., Shin M., Tamaki M., Terakawa H., Yoshiida N.;				
RA	"PURIFICATION, CHARACTERIZATION, CLONING, AND EXPRESSION OF A GLUTAMYL-ACID-SPECIFIC PROTEASE FROM BACILLUS LICHENIFORMIS ATCC 14580."				
RL	J. Biol. Chem. 267:23782-23788(1992).				
RN	[2]				
RP	SEQUENCE 95-316.				
RX	MEDLINE=92155199; PubMed=1346764;				
RA	Svendsen I., Breddam K.;				
RT	"Isolation and amino acid sequence of a glutamic acid specific endopeptidase from <i>Bacillus licheniformis</i> ."				
RL	Eur. J. Biochem. 204:165-171(1992).				
CC	-1- FUNCTION: SPECIFIC FOR HYDROLYSIS OF PEPTIDES BONDS ON THE CARBOXYL SIDE OF ACIDIC AMINO ACID RESIDUES, WITH A STRONG PREFERENCE FOR GLU.				
CC	-1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE: ASP-1-XAA, GLU-1-XAA.				
CC	-1- SUBCELLULAR LOCATION: SECRETED.				
CC	-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2B; ALSO KNOWN AS THE V8 FAMILY.				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.ebi.ac.uk/announce/ or send an email to license@ebi.ac.uk).				
CC	CC EMBL: D10060; BA00949.1; - .				
DR	DR PIR: S23078; S23078.				
DR	DR MEROPS: S01_271; - .				
DR	DR InterPro: IPR00126; Ser_proteas_V8.				
DR	DR InterPro: IPR001254; trypsin.				
DR	DR PF00099; trypsin; 1.				
DR	DR PRINTS: PRO0839; v8protease.				
DR	DR SMART: SM00020; tryp_spcl.				
DR	DR				

Result No.	Score	Query Match	Length	DB ID	Description
1	194	9.1	316	A45134	endopeptidase (EC metalloproteinase
2	135.5	6.4	313	A35122	coagulation factor
3	120.5	5.7	482	EXRT	pancreatic elastase
4	119	5.6	269	A26823	hypothetical protein
5	117	5.5	522	T29767	pancreatic elastase
6	115.5	5.4	269	B26823	pancreatic elastase
7	115.5	5.4	271	A25528	pancreatic elastase
8	115	5.4	258	S70439	pancreatic elastase
9	115	5.4	267	A56615	probable pancreatic
10	114.5	5.4	238	TRWV5Y	trypsin-like protease
11	114	5.3	246	DBHU	complement factor
12	113	5.3	266	EURL1	pancreatic elastase
13	113	5.3	266	ELPG	pancreatic elastase
14	112.5	5.3	259	TRSMG	trypsin (EC 3.4.21)
15	112.5	5.3	273	E85765	hypothetical protein
16	112.5	5.3	273	H64915	hypothetical protein
17	111.5	5.2	488	EXHU	coagulation factor
18	111	5.2	761	JCS759	brain-specific serine proteinase
19	110	5.2	269	C26823	pancreatic elastase
20	110	5.2	492	EXBO	coagulation factor
21	110	5.2	1582	T15308	hypothetical protein
22	109.5	5.1	405	T35117	probable secreted
23	108	5.1	583	A29154	complement factor
24	108	5.1	786	A77547	serine proteinase
25	108	5.1	1047	A55617	masquerade precursors
26	107.5	5.0	236	A26866	T-cell suppressor
27	107.5	5.0	686	A59271	Ra-reactive factor
28	105	4.9	274	S40004	trypsin-related protein
29	103.5	4.9	94	ELRT2	pancreatic elastase

Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2001, 16:17:55 : Search time 48.03 Seconds
(without alignments)
621,704 Million cell updates/sec

Title: US-09-072-384-15

Perfect score: 2131

Sequence: 1 MAGIPGULLFLLCAVGQ IKGNYLDCREGDTVFPFGSN 392

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5.

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing First 45 summaries

Database : PIR68:
1: Pir1:
2: Pir2:
3: Pir3:
4: Pir4:

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	194	9.1	316	A45134	endopeptidase (EC metalloproteinase
2	135.5	6.4	313	A35122	coagulation factor
3	120.5	5.7	482	EXRT	pancreatic elastase
4	119	5.6	269	A26823	hypothetical protein
5	117	5.5	522	T29767	pancreatic elastase
6	115.5	5.4	269	B26823	pancreatic elastase
7	115.5	5.4	271	A25528	pancreatic elastase
8	115	5.4	258	S70439	pancreatic elastase
9	115	5.4	267	A56615	probable pancreatic
10	114.5	5.4	238	TRWV5Y	trypsin-like protease
11	114	5.3	246	DBHU	complement factor
12	113	5.3	266	EURL1	pancreatic elastase
13	113	5.3	266	ELPG	pancreatic elastase
14	112.5	5.3	259	TRSMG	trypsin (EC 3.4.21)
15	112.5	5.3	273	E85765	hypothetical protein
16	112.5	5.3	273	H64915	hypothetical protein
17	111.5	5.2	488	EXHU	coagulation factor
18	111	5.2	761	JCS759	brain-specific serine proteinase
19	110	5.2	269	C26823	pancreatic elastase
20	110	5.2	492	EXBO	coagulation factor
21	110	5.2	1582	T15308	hypothetical protein
22	109.5	5.1	405	T35117	probable secreted
23	108	5.1	583	A29154	complement factor
24	108	5.1	786	A77547	serine proteinase
25	108	5.1	1047	A55617	masquerade precursors
26	107.5	5.0	236	A26866	T-cell suppressor
27	107.5	5.0	686	A59271	Ra-reactive factor
28	105	4.9	274	S40004	trypsin-related protein
29	103.5	4.9	94	ELRT2	pancreatic elastase

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing First 45 summaries

Database : PIR68:
1: Pir1:
2: Pir2:
3: Pir3:
4: Pir4:

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

ALIGNMENTS

RESULT	1
A45134	endopeptidase (EC 3.4.21), glutamate-specific - <i>Bacillus licheniformis</i>
C; Species: <i>Bacillus licheniformis</i>	
C; Date: 10-Jun-1993 #sequence-revision 18-Nov-1994 #text_change 15-Oct-1999	
C; Accession: A45134; S23078	
R; Kakudo, S.; Kikuchi, N.; Kitadokoro, K.; Fujiwara, T.; Nakamura, E.; Okamoto, H.; Sjövall, J.; Biol. Chem. 267, 23782-23786, 1992.	
A; Title: Purification, characterization, cloning, and expression of a glutamic acid specific endopeptidase	
A; Reference number: A45134; MUID: 93054737	
A; Accession: A45134	
A; Status: preliminary	
A; Molecule type: DNA	
A; Residues: 1-316 <RAK>	
A; Cross-references: GB:D1060; NID:9216263; PID:BAA00949.1; PID:d1001415; PID:921626	
A; Experimental source: ATCC 14580	
A; Note: sequence extracted from NCBI backbone (NCBIN:118784, NCBIP:118785)	
R; Svendsen, I.; Breddam, K.	
Bur. J. Biochem. 204, 165-171, 1992	
A; Title: Isolation and amino acid sequence of a glutamic acid specific endopeptidase	
A; Reference number: S23078	
A; Accession: S23078	
A; Status: preliminary	
A; Molecule type: protein	
A; Residues: 95-316 <SEVE>	
C; Keywords: hydrolase	
Query Match 9.1%; Score 194; DB 2; Length 316; Best Local Similarity 24.0%; Pred. No. 1.4e-08; Indels 64; Gaps 15; Matches 79; Conservative 43; Mismatches 143;	
Qy 63 SSGCPQCHKGKGPL--PTYEBAKQYLSYETLYANGSRTEHQVGYIYLSSSDGAQHRSGL 119	
Db 28 AQAAQSPHH-TPVSSSPSY-KAETSVTYDP----NIKSDQYGLYSKAFTGTGVNETKE 79	
Query Match 9.1%; Score 194; DB 2; Length 316; Best Local Similarity 24.0%; Pred. No. 1.4e-08; Indels 64; Gaps 15; Matches 79; Conservative 43; Mismatches 143;	
Qy 120 SSGCPQCHKGKGPL--PTYEBAKQYLSYETLYANGSRTEHQVGYIYLSSSDGAQHRSGL 119	
Db 80 KAEKKSPAKAPY--SIKSVGSDDRTRVNTTAPYRATWHISSIGSTGWMIGPKTV 136	
Db 171-TAAHETDCKT-YVKGTOKLRLVGFLKPKEFDGGRGANDTSAMPEQMKFOWIRVKRTHV 229	
Qy 171-TAAHETDCKT-YVKGTOKLRLVGFLKPKEFDGGRGANDTSAMPEQMKFOWIRVKRTHV 229	
Db 137 ATAGHCYDTSQSGFAGTATVSPG----RNGTS----YPGSVSK3TRYFI 178	
Qy 230 PKGWIKGNANDIGMDYDALLELKPHKRPMKIGVSSPAQLPGGRHFSGYNDRPGN 289	
Db 179 PSGWRSGNTN----YDYGAETLESEPTGNTVGYFSGSYTTSSLVGTIVTISGPDKTAG 233	
Qy 290 LVRFCDDYQQDCAQFGASGGVVYRMWKKRQOKWERKII---GIFSGH 343	
Db 234 TQWQHSGPIA SETKQIYAM-DTYYGGSGSPFPOSSRTNCSCPCLSLAVHTNGVYGG- 291	

F:103/Modified site: erythro-beta-hydroxyaspartic acid (Asp) #status predicted
 F:187/Binding site: carbohydrate (Asn) (covalent) #status experimental
 F:208/Binding site: carbohydrate (thr) (covalent) #status predicted
 F:218-/332/Cleavage site: carbohydrate (Asn) (covalent) #status predicted
 F:231-/340/Active site: Arg-Ile (coagulation factor IXa, coagulation factor VIIa) #status predicted
 F:274-/320,417/Active site: His, Asp, Ser #status predicted

Query Match 5.74; Score 120.5; DB 1; Length 482;

Best Local Similarity 32.3%; Pred. No. 0..03; 23; Indels 25; Gaps 4;

Matches 31; Conservative 17; Mismatches 23; Gaps 4;

Db 160 CTGTLVAAEKAIVTAACIHD**T**YKRGFLPKPFDGGGRANDSISAMPQMKF 219

Db 259 CGGTILNEFY**T**TAACLHQKRF---KVRGDLNTEQEDGEWHE-WDMIIKHNF 312

Db 313 Q---RDY-----DFDIAMRLRKTP 329

RESULT 4

A26823

pancreatic elastase II (EC 3.4.21.71) precursor - pig

Ci:Alternate names: pancreaticopeptidase E

Ci:Species: Sus scrofa domestica (domestic Pig)

CDte: 16-Aug-1988 #sequence_change 22-Jun-1999

Ci:Accession: A26823

Ri:Kawashima, I.; Tani, T.; Shimoda, K.; Takiguchi, Y.

DNA 6, 163-172, 1987

A:Title: Characterization of pancreatic elastase II cDNAs: two elastase II mRNAs are exp

A:Reference number: A90958; MUID:87217962

A:Accession: A26823

A:Residues: 1-269 <KAW>

A:Cross-references: GB:MI16651; NID:9164441; PID:AAA31027.1; PID:9164442

Ci:Superfamily: trypsin; trypsin homology

Ci:Keywords: hydrolase; serine proteinase

F:1-16/Domain: signal sequence #status predicted <SIG>

F:17-28/Domain: protease II #status predicted <PRO>

F:29-289/Product: elastase II #status predicted <MAT>

F:29-362/Domain: trypsin homology <TRY>

F:73,121,216/Active site: His, Asp, Ser #status predicted

Query Match 5.68; Score 119; DB 2; Length 269;

Best Local Similarity 27.5%; Pred. No. 0..02; 28; Mismatches 64; Indels 66; Gaps 14;

Matches 60; Conservative 28; Mismatches 64; Indels 66; Gaps 14;

Db 128 RQIGYDSRFISFGKD**T**LNNPFESTVSKL-STG-----CTGTLVAEH**T**HVLTAHCTHDGK 181

Db 28 RVGGEDARPN -----SWPQVSLOYDSSGQWRHTCGTGLVOSWLVTAHCTSSR 79

Db 182 TY--VKGTO**T**LRGVFLPKPFDGGGRANDSISAMPQMKF**T**WIRVKRTHVKGWIGNAN 239

Db 80 TYRVVLGRHSL-----STNEPSLA-----KVSKVLLVHQDW---NN 114

Db 240 DGMDDYDALL**T**EKKLPK HKRKFMKIGSPPAKQ-LPG-----GRIHSQYND--R 286

Db 115 QLSNGNDIALKLKASPLSVLTDKIQLGCLPAAGTLEPNNYCYXTGWMRLQNTGASFDILQ 174

Db 287 PGNLVYRFCDYKDETDLYQOCDAQFGASGSQYYRM 324

Db 175 QEQLL-----YATC-SKPFWMGSTVKNM 200

Db 175 QEQLL-----YATC-SKPFWMGSTVKNM 200

Db 175 QEQLL-----YATC-SKPFWMGSTVKNM 200

Query Match 5.58; Score 117; DB 2; Length 522;

Best Local Similarity 20.4%; Pred. No. 0..065; 56; Mismatches 153; Indels 162; Gaps 21;

Db 29 KPTWPAYRLPVVLPQSTNLAKPDPFGAEAKLEVSSS--CGPQCHKGTPLYEEAKQ-- 83

Db 26 KPSNKASSAPSRLRKKSNSNPNGKTARSVSKYSSAIPASPTVKEVPPVTEIEKKEEK 85

Db 86 PENQKKELAEKKL---DRTQDDKEYKEAESALGVVIKEDKAPAMDG1VEDFGPCCEF 141

Db 138 SIFGKD-----ELLNYFESTSV-----KLSLTGTTGTLVAK 168

Db 142 PFLKLDEMILWLRIFPSAKYNGRRAQSAPWSYTFYLSKDEESATCTGTIVSPR 201

Db 169 HVLTAAHCI----HDGK-----TYVKGTOKL--RVGFLKPKFKDGRGA 206

Db 202 HILIATHICFAQGNRQDGWNLIETDERSNCNDQDDYVITNOEFLKRVFLSNK----- 253

Db 207 NDSTSAMEQMKFQWIVTRKTHVPGIKGNANDIGHDY---DYALLEL----- 252

Db 254 -KGISRPPEKITHLVACTKRT-----ANRTRKIPPOQYTDDEFAITHYLEELTFSSNVQ 305

Db 253 -----KKPIKR---KFMKIGVSPAKOLPGRIFSGYDNDPQY----- 292

Db 306 SVCVADDETQPNDKLSLEYFGGLNPNSD-----INQNGVDN-TGQLRYKEIVFRSH 357

Db 293 -----RFCDVYKDETVD-----UYQOQCDAQ-----PGASGSVYVRMMKROQQKWE 333

Db 358 PNEIYFFQARDITDTKTVACVVLKILTKTOASLNLSKGDSGGGIADV-----KGK 411

Db 412 KTGIGLSQTSCQKRRSGNEMLYSVGFYKQNI---CKYTGIC 453

RESULT 6

B26823

pancreatic elastase II (EC 3.4.21.71) A precursor - human

Ci:Species: Homo sapiens (man)

Ci:Date: 16-Aug-1988 #sequence_revision 16-Aug-1988 #text_change 20-Jun-2000

Ri:Kawashima, I.; Tani, T.; Shimoda, K.; Takiguchi, Y.

DNA 6, 163-172, 1987

A:Title: Characterization of pancreatic elastase 2 determined by sequence anal

A:Reference number: A90958; MUID:87217962

A:Accession: B26823

A:Molecule type: mRNA

A:Residues: 1-269 <KAW>

A:Cross-references: GB:MI16652; NID:9182057; PID:AAA52380.1; PID:9182058

Ri:Fletcher, T.S.; Shen, W.P.; Largman, C.

Biochemistry 26, 7256-7261, 1987

A:Title: Primary structure of human pancreatic elastase 2 determined by sequence anal

A:Reference number: A27432; MUID:88107669

A:Molecule type: mRNA

A:Residues: 1-269 <FILE>

RESULT 5

T29767

hypothetical protein ZC581.6 - Caenorhabditis elegans

Ci:Species: Caenorhabditis elegans

Ci:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

Ci:Accession: T29767

Ri:Waterston, R.; Gattung, S.; Le, T.T.

A;Cross-references: GB:MI6631; NID:g182022; PIDN:AAA52374_1; PID:g182023
 R; Shirasu, Y.; Yoshida, H.; Matsuki, S.; Ikeda, N.; Shimada, Y.; Ozawa, T.
 J; Biochem. 102, 1555-1563, 1987
 A;Title: Molecular cloning and expression in Escherichia coli of a cDNA encoding human F
 A;Reference number: A41431; MUID:88198076
 A;Accession: A41431
 A;Molecule type: mRNA
 A;Residues: 1-201, V, 203-269 <SH1>
 A;Note: the authors translated the codon GRS for residue 202 as Cys
 R; Moullé, M.; Michon, T.; Kerfrel, B.; Chapus, C.
 FEBS Lett. 261, 179-183, 1990
 A;Title: Further studies on the human pancreatic binary complexes involving procarboxypeptidase
 A;Reference number: S08253; MUID:90169111
 A;Accession: S34491
 A;Molecule type: protein
 A;Residues: X, 18-50 <NOU>
 C;Genetics:
 A;Gene: GDB:ELA1
 A;Cross-references: GDB:119866; OMIM:130120
 A;Map position: 12pter-12qter
 C;Superfamily: trypsin; trypsin homology
 C;Keywords: hydrolase; pancreas; serine proteinase
 F;1-16/Domain: signal; sequence #status predicted <SIG>
 F;17-28/Domain: propeptide #status predicted <PRO>
 F;29-269/Product: pancreatic elastase LIA #status predicted <MAT>
 F;29-262/Domain: trypsin homology <TRY>
 F;73,121-216/Active site: His, Asp, Ser #status predicted

Query Match 5.4% Score 115.5; DB 2; Length 269;
 Best Local Similarity 28.6%; Pred. No. 0.038; Gaps 11;
 Matches 50; Conservative 21; Mismatches 51; Indels 53; Gaps 11;

QY 147 NYPESTSYKLSTG-----CTGTLYVAE~~H~~VLTAAHC~~I~~HDGKTYVKG~~T~~QKLRYGFLKPKF 200
 DB 39 SWPQVSLOYSSNGKWHTCGS~~L~~ANSVLTAAHC~~S~~RRY-----RVL-----85

QY 201 DGGRG---ANDTSAMPEQM~~K~~QWIRV~~A~~RT~~P~~GWIKGNANDIGMDYD~~Y~~ALLEL~~K~~PKF- 255
 DB 86 -GRHNL~~Y~~VAEGS~~L~~SA-----VSV~~T~~SKIVYHKDW--NSHQIS~~K~~GND~~I~~ALKLN~~P~~ 131

QY 256 HKREMKIGVSPPAKQ-LPG-----GRIHFGSYND-RPGNLY--YRFC 295
 DB 132 SLTDKIQLACPPGTLDPNNYPCV~~T~~GWGRLLQTNGAVPDVLOGRLLVVDATC 186

RESULT 7
 Pancreatic elastase II (EC 3.4.21.71) precursor - mouse
 C;Species: Mus musculus (house mouse)
 C;Accession: A25528
 R; Stevenson, B.J.; Hagenbuechle, O.; Wellauer, P.K.
 R; Nucleic Acids Res. 14, 8307-8330, 1986
 A;Title: Sequence organisation and transcriptional regulation of the mouse elastase II a
 A;Reference number: A93646; MUID:87066713
 A;Accession: A25528
 A;Molecule type: mRNA
 A;Residues: 1-271 <STE>
 A;Cross-references: GB:X04573; NID:g50825; PIDN:CAA28242_1; PID:g50826
 C;Superfamily: trypsin; trypsin homology
 C;Keywords: hydrolase; serine proteinase
 F;1-10/Domain: signal sequence #status predicted <SIG>
 F;31-271/Product: pancreatic elastase II #status predicted <MAT>
 F;31-264/Domain: trypsin homology <TRY>
 F;75,123,218/Active site: His, Asp, Ser #status predicted

Query Match 5.4% Score 115.5; DB 2; Length 271;
 Best Local Similarity 25.0%; Pred. No. 0.039; Gaps 18;
 Matches 67; Conservative 38; Mismatches 80; Indels 83; Gaps 18;

QY 148 YPFSTSVK-LSTG-----CTGTLYVAE~~H~~VLTAAHC~~I~~HDGKTY- VKGTQKLRYGFLKPKF 199
 DB 42 WPQVSLOYSSNGKWHTCGS~~L~~ANSVLTAAHC~~S~~RRY-----RVL-----92

QY 200 KDGGRGANDTSAMPEQM~~K~~QWIRV~~T~~HPKG~~N~~ANDIGMDYD~~Y~~ALLEL-KKPHK 258
 DB 93 -SNPGAGSAA-----VQVSKLVHQ~~R~~W--NSQNVNGYD~~T~~ALIKLASPV~~T~~IS 136

QY 259 KEMKIGVSPPAKQ-LPGGRI-HFSGY-----DRPGNLY--YRFC-----295
 DB 137 KNIQAC~~L~~PPAGT~~T~~LP~~R~~PRN~~V~~CYVGWGLLQTNGNSPDT~~R~~QGR~~L~~VVDATC~~S~~ASWNGS 196

QY 296 DVKDETY---DLIYQCDADAQPAGSG~~S~~GYVVR~~M~~KROOKWERK~~I~~IGFSGHQWDMNGS 351
 DB 197 SVKSM~~M~~CAAGGDGT~~S~~SCN---GDSGPINCRASNGO--WQ--VHGTVSF~~G~~SSLGC~~N~~Y 247

QY 352 PQDENVA~~R~~ITPLK~~Y~~AQICYWK~~N~~YL~~D~~ 379
 DB 248 PRKF~~S~~V~~F~~TRV~~S~~-----NYID 262

RESULT 8
 S70439
 Pancreatic elastase I (allele HEL-16) probable splice form I - human
 C;Species: Homo sapiens (man)
 C;Date: 03-Aug-1998 #sequence_revision 03-Aug-1998 #text_change 16-Dec-1998
 C;Accession: S70439
 R; Kawashima, I.; Tani, T.; Shimoda-Takano, K.; Ohmine, T.; Furukawa,
 DNA Seq. 2, 303-312, 1992
 A;Title: Genomic organization of the human homologue of the rat pancreatic elastase I
 A;Reference number: A56615; MUID:9233395
 A;Status: preliminary; not compared with conceptual translation
 A;Accession: S70439
 A;Molecule type: DNA
 A;Residues: 1-258 <KAW>
 C;Superfamily: trypsin; trypsin homology
 F;19-251/Domain: trypsin; trypsin homology <TRY>

Query Match 5.4%; Score 115; DB 2; Length 258;
 Best Local Similarity 26.9%; Pred. No. 0.04; Gaps 6;
 Matches 35; Conservative 23; Mismatches 38; Indels 34; Gaps 6;

QY 147 NYPESTSYKLSTG-----CTGTLYVAE~~H~~VLTAAHC~~I~~HDGKTY- VKGTQKLRYGFLKPKF 198
 DB 29 SWPQISL~~O~~YRSGGS~~W~~YTCGGT~~L~~IQRNQVMTAHC~~V~~YDQKTPRVA~~V~~ADHNL-----80

QY 199 FKDGGRGANDTSAMPEQM~~K~~QWIRV~~T~~HPKG~~N~~ANDIGMDYD~~Y~~ALLEL-KKPHK 257
 DB 81 -----SQNDGTE-----QIVVSQKV~~V~~HPW--NSDNVAA~~G~~YD~~T~~LLR~~A~~Q~~S~~VT~~L~~ 123

QY 258 RKFMKIGSP 267
 DB 124 NSTVQLGVLP 133

RESULT 9
 A56615
 probable pancreatic elastase (EC 3.4.21.36) pseudogene - human
 N;Alternate names: pancreatic elastase I homolog; pancreatic elastase I allele HEL1-1
 C;Species: Homo sapiens (man)
 C;Date: 11-Aug-1995 #sequence_revision 17-Aug-1995 #text_change 14-Aug-1998
 C;Accession: A56615; S70440
 R; Kawashima, I.; Tani, T.; Mita-Honjo, K.; Shimoda-Takano, K.; Ohmine, T.; Furukawa,
 DNA Seq. 2, 303-312, 1992
 A;Title: Genomic organization of the human homologue of the rat pancreatic elastase I
 A;Reference number: A56615; MUID:9233395
 A;Molecule type: DNA
 A;Residues: 1-267 <KAW>
 A;Cross-references: EMBL:X62259; NID:931246; EMBL:X62258; GB:S40923; NID:931247; EMBL:
 5; GB:SA0856; NID:931251; EMBL:X62256; GB:S40857; NID:93152; EMBL:62257; GB:SA0859;
 A;Note: sequence extracted from NCBI backbone (NCBIN:109315, NCBIN:109317, NCBIN:1093

C;Comment: This apparently silent human homolog of pancreatic elastase I is a single-copy protein in some other tissue.

C;Generics:

A;Gene: GDB:ELA1

A;Map Position: 12

C;Keywords: hydrolase; serine proteinase

Query Match 5.4%; Score 115; DB 4; Length 267;
Best Local Similarity 26.9%; Pred. No. 0.042; Indels 34; Gaps 6;
Matches 35; Conservative 23; Mismatches 38; Amino acid sequence of human D of the alternative complement pathway.

Qy 147 NYPFSTSVKLSLTG-----CTGTLVAEKHVLTAACIHHDGKTY--VKGTKQLRVLGFKLKP 198
Db 38 SWPSQISLQLQRGGSSWHTCGTLLIRQNWTAAHCYDQNTFRVAGDHNH----- 89
Qy 199 FKDGGRGANDSTSAMPQMKKFWIRVKRTHPKGWIKGNANDIGMDYALLEL-KKPHK 257
Db 90 -----SQNDTE-----QVSVQKIVVHPW---NSDVAAGCYDIALURLAQSVTL 132

Qy 258 RKFMKIGVSP 267
Db 133 NSYVOLGVLP 142

RESULT 10

TRWV5Y

trypsin-like proteinase (BC 3.4.21.-) 5G1 precursor - yellow fever mosquito (fragment)

C;Species: Aedes aegypti (yellow fever mosquito)

C;Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 18-Jun-1999

A;Accession: S19891

R;Kalhok, S.; Tabak, L.M.; Prosser, D.E.; Downe, A.E.R.; White, B.N.
submitted to the EMBL Data Library, February 1992

A;Description: Isolation, sequencing and characterization of 2 cDNA clones coding for trypsin-like proteinase (BC 3.4.21.-) 5G1 precursor - yellow fever mosquito (fragment)

A;Reference number: S19890

A;Accession: S19891

A;Molecule type: mRNA

A;Residues: 1-238 <HLA>

A;Cross-references: EMBL:X64363; NID:g5563; PID:CAA45715.1; PID:95564

C;Superfamily: trypsin; trypsin homology

C;Keywords: hydrolase; insect midgut; protein digestion; serine proteinase

F;1-11/Domain: signal sequence and propeptide (fragment) #status predicted

F;12-232/Domain: trypsin homology <TRY>

F;38-54, 162-178, 189-213/disulfide bonds; #status predicted

F;53,97,193/Active site: His, Asp, Ser #status predicted

Query Match 5.4%; Score 114.5; DB 1; Length 238;
Best Local Similarity 29.4%; Pred. No. 0.04; Indels 44; Gaps 7;
Matches 42; Conservative 22; Mismatches 44; Amino acid sequence of human D of the alternative complement pathway.

Qy 149 PRSTS --VRLSTGCTGTVAEKHVLTAACIHHDGKTYVKTQKLRYFLPKFKDGGRG 205
Db 24 PFQVSLSGVGSSHFCGSILSERWVNTAGHQAASGOTNL---QVHIG---SSQHASSG-- 75

Qy 206 ANDSTSAMPQMKFQTRVKRTHPKGWIKGNANDIGMDYALLELKKPKRKEMKIG 264

Db 76 -----QI.KVKVNRHPK-----YDEVTVDYDFALLEEETVTFSDSCAP 115

Qy 265 VSPPAKQLP---GGRHFSGDN 284
Db 116 VKLPOKDTPVNEGTCLOVSGGN 138

RESULT 12

ELRT1

pancreatic elastase (EC 3.4.21.36) I precursor - rat

C;Species: Rattus norvegicus (Norway rat)

C;Date: 18-Aug-1982 #sequence_revision 18-Aug-1982 #text_change 18-Jun-1999

C;Accession: A00560; A20534

R;MacDonald, R.J.; Swift, G.H.; Quinto, C.; Swain, W.; Pictet, R.L.; Nikovits, W.; Ru

Biochemistry 21, 1453-1463, 1982

A;Title: Primary structure of two distinct rat pancreatic preproelastases determined

A;Reference number: A00560; MUID:82182967

A;Accession: A00560

A;Molecule type: mRNA

Query Match 5.3%; Score 114; DB 1; Length 246;

Best Local Similarity 28.4%; Pred. No. 0.046; Indels 43; Gaps 9;

Qy 149 PEFTSYKLSTG--CTGTLVAEKHVLTAACIHHDGKTYVKTQKLRYFLPKFKDGG 203
Db 31 PYMASVQLNGAHLCAAGVLAERWVLSAHCLEDAADGKVQL----- 72

Qy 204 RGANDSTSAMPQMKFQTRVKRTHPKGWIKGNANDIGMDYALLELKKPKRKEMKIG 263
Db 73 LGAHSLSQPEPSKRLYDLRA---YB---HPDSQDTIDHLILQSE-----KA 117

Qy 264 GVSPPAKQLPGRHFGESGYDND-RPGNLYVRFCDV 297
Db 118 TLGPAPVPLPQRV-----DRDVAPGTL-----CDV 143

RESULT 11

DBHU

complement factor D (EC 3.4.21.46) precursor [validated] - human (fragment)

C;Alternate names: adipoin; C3 convertase activator

C;Species: Homo sapiens (man)

C;Date: 28-Aug-1985 #sequence_revision 31-Dec-1992 #text_change 08-Dec-2000

C;Accession: A40197; A00936; A60571; S6645

R;White, R.T.; Damm, D.; Hancock, N.; Rosen, B.S.; Lowell, B.B.; Usher, P.; Flier, J.S.;

J. Biol. Chem. 267, 9210-9213, 1992
A;Title: Human adipoin is identical to complement factor D and is expressed at high 1

A;Reference number: A40197; MUID:92220520

A;Accession: A40197

A;Molecule type: mRNA

A;Residues: 1-246 <W11>

A;Cross-references: GB:MB4526

R;Niemann, M.A.; Bhown, A.S.; Bennett, J.C.; Volanakis, J.E.

Biochemistry 23, 2482-2486, 1984

A;Title: Amino acid sequence of human D of the alternative complement pathway.

A;Reference number: A00936; MUID:85000441

A;Accession: A00936

A;Molecule type: protein

A;Residues: 19-44, 'G', '46-51, 'Q', '53-75, 'P', '80-83, 'XXXIT', '90-172, '86-91, '185-23

A;Note: a few residues were assigned from the previously published sequence of Reid et al.

R;Miyata, T.; Oda, O.; Inagi, R.; Sugiyama, S.; Miyama, K.; Nakashima, I.; Mol. Immunol. 27, 637-644, 1990

A;Title: Molecular and functional identification and purification of complement compo-

A;Reference number: A60571; MUID:90370044

A;Accession: A60571

A;Molecule type: protein

A;Residues: 19-20, 'XX', '23-27, 'XX', '30-31, 'XX', '34, 'X', '36-40 <MY>

R;Balke, N.; Holzkamp, U.; Hoerl, W.H.; Tschesche, H.

FEBS Lett. 371, 300-302, 1995

A;Title: Inhibition of degranulation of human polymorphonuclear leukocytes by comple-

A;Reference number: S66645; MUID:96013156

A;Accession: S66645

A;Status: preliminary

A;Molecule type: protein

A;Residues: 19-44, 'C', '46-48 <BAL>

C;Comment: Factor D cleaves factor B when the latter is complexed with factor C3b, ac

C;Genetics:

A;Gene: GDB:DF

A;Cross-references: GDB:136645; OMIM:134350

A;Map position: Xpter-Xqter

C;Superfamily: trypsin homology

C;Keywords: complement alternate pathway; hydrolase; plasma; serine proteinase

F;1-18/Domain: signal sequence #status predicted <SIG>

F;19-246/Product: complement factor D (fragment) #status experimental

F;19-241/Domain: trypsin homology <TRY>

F;44-60,140-207,172-188,191-222/Disulfide bonds: #status predicted

F;59,105,201/Active site: His, Asp, Ser #status predicted

A;Residues: 1-266 <MAC>
 A;Cross-references: GB:V01234; NID:956088; PIDN:CAA24544.1; PID:956089
 R;Largman, C.
 Biochemistry 22, 3763-3770, 1983
 A;Title: Isolation and characterization of rat pancreatic elastase.
 A;Reference number: A20534; MUID:84000385
 A;Accession: A20534
 A;Molecule type: protein
 A;Residues: 17-37, 'X', 39-45 <LR>
 C;Superfamily: trypsin; trypsin homology
 C;Keywords: hydrolase; pancreas; serine proteinase; zymogen
 F;1-16/Domain: signal sequence #status predicted <SIG>
 F;17-26/Domain: activation peptide #status predicted <APT>
 F;27-29/Domain: elastase I #status predicted <MP1>
 F;27-29/Domain: trypsin homology <TRY>
 F;71,119,214/Active site: His, ASP, Ser #status predicted

Query Match 5.3%; Score 113; DB 1; Length 266;
 Best Local Similarity 27.3%; Pred. No. 0.061; Indels 34; Gaps 6;
 Matches 36; Conservative 21; Mismatches 41; Indels 34; Gaps 6;

QY 147 NYPFSTSVKLSTG -----CTGTLLVAEKHVLTAAHCINHDGKTY -VKGTOKLRLVGFELPK 198
 Db 37 SWPQISLQYRSGSSWAHCGGTLLRQNWWMTAAHCYDRELTRVYVGEHNL----- 88

Query Match 5.3%; Score 113; DB 1; Length 266;
 Best Local Similarity 27.3%; Pred. No. 0.061; Indels 34; Gaps 6;
 Matches 36; Conservative 21; Mismatches 41; Indels 34; Gaps 6;

QY 147 NYPFSTSVKLSTG -----CTGTLLVAEKHVLTAAHCINHDGKTY -VKGTOKLRLVGFELPK 198
 Db 37 SWPQISLQYRSGSSWAHCGGTLLRQNWWMTAAHCYDRELTRVYVGEHNL----- 88

Query Match 5.3%; Score 113; DB 1; Length 266;
 Best Local Similarity 26.78%; Pred. No. 0.061; Indels 36; Gaps 7;
 Matches 39; Conservative 23; Mismatches 48; Indels 36; Gaps 7;

QY 128 RQIYGDSRFSTFKDFLLNPFFTSVVLSTG ---CTGTLLVAEKHVLTAAHCINHDGKT 182
 Db 26 RIVVGAEARNSNPQSISQY -----LGGGSWYHTCGTLLRNRNWMTAAVCVSSQMT 78

Query Match 5.3%; Score 113; DB 1; Length 266;
 Best Local Similarity 26.78%; Pred. No. 0.061; Indels 36; Gaps 7;
 Matches 39; Conservative 23; Mismatches 48; Indels 36; Gaps 7;

QY 183 YVKGTOKLRLVGFLPKPKFDGGRGANDSTSAMPBOMKFQWIRVTHRTHPKGWIGKGNANDG 242
 Db 79 F-----RIVVG -----DHNLSONDGT-----QYVSQVKIMVHPTW --NSNNVA 115

Query Match 5.3%; Score 113; DB 1; Length 266;
 Best Local Similarity 26.78%; Pred. No. 0.061; Indels 36; Gaps 7;
 Matches 39; Conservative 23; Mismatches 48; Indels 36; Gaps 7;

QY 243 MDYDYLLEL-KKPHKREPKMKIGVSP 267
 Db 116 AGYDIALRLQASVTLANNVQLAVLP 141

RESULT 13
 ELPG
 Pancreatic elastase (EC 3.4.21.36) I precursor - pig
 C;Species: Sus scrofa domestica (domestic pig)
 C;Date: 24-Apr-1984 #sequence revision 30-Sep-1990 #text_change 16-Jun-2000
 C;Accession: JS0013; A26777; A10061; A00959
 R;Shirasu, Y.; Yoshida, H.; Miyayama, T.; Matsuki, S.; Tanaka, J.I.; Ikenaga, H.
 J. Biochem. 95, 1707-1712, 1986
 A;Title: Isolation and expression in Escherichia coli of a cDNA clone encoding porcine pancreatic elastase. A;Reference number: A92005; MUID:86304235
 A;Accession: JS0013
 A;Molecule type: protein
 A;Residues: 1-266 <SH1>
 A;Cross-references: GB:X04'036; GB:D00070; GB:N00070; PIDN:BAA00118.1; PID:9217684
 R;Tani, T.; Kawashima, I.; Furukawa, H.; Ohmine, T.; Takiguchi, Y.
 J. Biochem. 101, 591-599, 1987
 A;Title: Characterization of a silent gene for human pancreatic elastase I: structure of A;Reference number: A26777; MUID:87250343
 A;Accession: A26777
 A;Molecule type: protein
 A;Residues: 1-125, 'G', 127-183, 'L', 185-266 <TAN>
 A;Cross-references: GB:D00160; NID:9217683; PIDN:CAA27670.1; PID:9217684
 R;Note: the authors translated the codon GGC for residue 58 as Gln, GGC for residue 126
 R;Shotton, D.M.; Hartley, B.S.
 Biochem. J. 131, 643-645, 1973
 A;Title: Evidence for the amino acid sequence of porcine pancreatic elastase.
 A;Reference number: A90267; MUID:73229121
 A;Accession: A10061
 A;Molecule type: protein
 A;Residues: 27-91, 'N', 93-203, 'N', 205-266 <SH0>
 R;Shotton, D.M.; Hartley, B.S.
 Biochemistry 22, 811-816, 1970
 A;Title: Three-dimensional structure of tosyl-elastase.
 A;Reference number: A93160; MUID:70114044
 A;Content: annotation: X-ray crystallography, 3.5 angstroms; active site
 C;Superfamily: trypsin; trypsin homology
 C;Keywords: hydrolase; pancreas; serine proteinase; zymogen
 F;1-16/Domain: signal sequence #status predicted <SIG>

Query Match 5.3%; Score 112.5%; DB 1; Length 259;
 Best Local Similarity 26.0%; Pred. No. 0.065; Indels 37; Gaps 5;
 Matches 39; Conservative 21; Mismatches 53; Indels 37; Gaps 5;

QY 148 YPFSTSVKLSTGCTGTLVAEKHVLTAAHCINHDGKTYVKGTOILRVGFELPKFKDGGRCAN 207
 Db 48 YPFSTSVKLSTGCTGTLVAEKHVLTAAHCINHDGKTYVKGTOILRVGFELPKFKDGGRCAN 207
 QY 208 DS---TSSAMPEQMKFWIRVTHRTHPKGWIGKGNANDGMDYALLEKKPHKREPKMKIG 264

Db 82 TSIATGGVVDLQSSAVKVRSTKVLOA---PGYNGTGGDWALIKLAQPINOPTLKIA 136
 C:Species: Escherichia coli
 C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
 C:Accession: E85765

Ri:Perna, N.R.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca, Nature 409, 529-533, 2001

A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7

A:Reference number: A85480; PMID:21074935; PMID:11206551

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-273 <STOP>

A:Cross-references: GB:AE005174; NID:912515576; PIDN:AA56585.1; GSPDB GN00145; UWGP:Z25

A:Experimental source: strain O157:H7, substrate EDL933

C:Genetics:

A:Gene: Z2592

Query Match 5.38; Score 112.5; DB 2; Length 273;
 Best Local Similarity 23.3%; Pred. No. 0 0.069;
 Matches 62; Conservative 35; Mismatches 112; Indels 57; Gaps 13;

Qy 101 VGYIILSSSGSCKAQHRDGSSCKSRRRQIIGYDSDPSRFGRDFLLNPFFESTSVKUSTG- 159
 Db 9 LGRALISLTSASFVPAKDVAKSANDEVSTLFGHDDRVPV--NDTQSPWMAVGQLETTAS 65

Qy 160 ---CTGTLYAEEKHVLTAACHTDGGKTYVKKGQFKQLRKGFLKKPKFKDGGR-----GANDS 209
 Db 66 GNNCTATLAPNLALTAGHCL---LTPPKGRADKAYALRFVSNKGWLRYDHIDIEGRVP 122

Qy 210 TSAMPEOMKFQWIRVKRTHVPKGWIGKGNANDIGMDYDALLELKKPKHKMKIGVSP-- 267
 Db 123 T---LGKRLKAD-----GDGWIVPPA---AAPNDFGLIVLRNPPS-----GITPLP 162

Qy 268 -----PAKQLPGRHIFGSDYIDRPGNLV-RFCDYLDIQQCDAQPGA 315
 Db 163 LFEQDKAALTAAALKAAAGRKVTOAGYQPDHDLTLYSHONCEVTGWAQTTSVMSHQCDTLPGD 222

Qy 316 SGGGVYRMWKRQQQKWERKLIGIFS 341

Db 223 SGSPML-----HTDDGMQ--LIGVQS 242

Search completed: December 2, 2001, 16:17:56
 Job time: 191 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2001, 16:14:40 ; Search time 35.78 Seconds
(without alignments)

246.542 Million cell updates/sec

Title: US-09-072-384-2
Perfect score: 2112
Sequence: 1 MAGIPGILFLFLCAVGQ IKGNYLDCREGDTVFLPGSN 3 92

Scoring table: BLOSUM62
Gapext 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA:
 1: /cggn2_6/pcdatad/2/1aa/5A_COMB.pep:
 2: /cggn2_6/pcdatad/2/1aa/5B_COMB.pep:
 3: /cggn2_6/pcdatad/2/1aa/6A_COMB.pep:
 4: /cggn2_6/pcdatad/2/1aa/6B_COMB.pep:
 5: /cggn2_6/pcdatad/2/1aa/PCTUS_COMB.pep:
 6: /cggn2_6/pcdatad/2/1aa/backfile1.pep:
 *
 *
 *

Pred. No. is the number of results Predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Match	Length	DB ID	Description
1	2104	99.6	392	4	US-09-072-384-2	Sequence 2, Appl1
2	2085	98.7	392	4	US-09-072-384-15	Sequence 15, Appl1
3	2044	96.8	383	4	US-09-072-384-18	Sequence 1, Appl1
4	165	7.8	222	1	US-08-090-048-1	Sequence 1, Appl1
5	165	7.8	222	2	US-08-092-550-1	Sequence 1, Appl1
6	165	7.8	222	2	US-07-922-651A-1	Sequence 1, Appl1
7	123	5.8	256	3	US-08-096-769-89	Sequence 89, Appl1
8	123	5.8	256	3	US-08-096-616-89	Sequence 89, Appl1
9	123	5.8	256	4	US-08-817-795-89	Sequence 89, Appl1
10	123	5.8	256	4	US-08-639-075A-89	Sequence 89, Appl1
11	123	5.8	256	4	US-09-012-431-89	Sequence 89, Appl1
12	123	5.8	256	4	US-09-032-215-32	Sequence 32, Appl1
13	123	5.8	256	4	US-09-012-612-89	Sequence 89, Appl1
14	123	5.8	256	4	US-08-096-613-89	Sequence 89, Appl1
15	123	5.8	256	5	PCT-US5-14442A-89	Sequence 2, Appl1
16	121.5	5.8	437	1	US-08-487-037-2	Sequence 59, Appl1
17	118	5.6	241	4	US-08-944-483-59	Sequence 3, Appl1
18	116.5	5.5	437	1	US-08-487-037-3	Sequence 3, Appl1
19	114	5.4	254	1	US-08-330-978-3	Sequence 3, Appl1
20	114	5.4	254	1	US-08-474-042-3	Sequence 3, Appl1
21	114	5.4	254	1	US-08-484-538-3	Sequence 3, Appl1
22	114	5.4	254	1	US-08-747-532-3	Sequence 3, Appl1
23	114	5.4	306	1	US-08-130-978-1	Sequence 1, Appl1
24	114	5.4	306	1	US-08-474-042-1	Sequence 1, Appl1
25	114	5.4	306	1	US-08-744-538-1	Sequence 1, Appl1
26	114	5.4	306	1	US-08-774-532-1	Sequence 1, Appl1
27	114	5.4	448	1	US-08-295-411-3	Sequence 1, Appl1

ALIGNMENTS

RESULT 1
US-09-072-384-2
; Sequence 2, Application US/09072384
; Patent No. 6153420

GENERAL INFORMATION:
 / APPLICANT: Sheppard, Paul O.
 / TITLE OF INVENTION: SERINE PROTEASE POLYPEPTIDES
 / NUMBER OF SEQUENCES: 18
 / CORRESPONDENCE ADDRESS:
 / ADDRESS: ZymoGenetics, Inc.
 / STREET: 1201 Eastlake Avenue East
 / CITY: Seattle
 / STATE: WA
 / COUNTRY: USA
 / ZIP: 98102

COMPUTER READABLE FORM:
 / MEDIUM TYPE: Diskette;
 / COMPUTER: IBM Compatible
 / OPERATING SYSTEM: DOS
 / SOFTWARE: FastSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:
 / APPLICATION NUMBER: US/09/072, 384
 / FILING DATE:
 / CLASSIFICATION:
 / PRIORITY APPLICATION DATA:
 / APPLICATION NUMBER:
 / FILING DATE:
 / ATTORNEY/AGENT INFORMATION:
 / NAME: Parker, Gary E.
 / REGISTRATION NUMBER: 31, 648
 / REFERENCE/DOCKET NUMBER: 91-16C1
 / TELECOMMUNICATION INFORMATION:
 / TELEPHONE: 206-442-6673
 / TELEX: 206-442-6678

INFORMATION FOR SEQ ID NO: 2:
 / SEQUENCE CHARACTERISTICS:
 / LENGTH: 392 amino acids
 / TYPE: amino acid
 / STRANDEDNESS: single
 / TOPOLOGY: linear
 / MOLECULE TYPE: protein
 / FRAGMENT TYPE: internal
 / FEATURE:
 / NAME/KEY: Signal Sequence

OTHER INFORMATION:
 / LOCATION: 1...19
 / OTHER INFORMATION:
 / US-09-072-384-2

Query Match 99.6%; Score 2104; DB 4; Length 392;
 Best Local Similarity 100.0%; Pred. No. 6.e-21.9;
 Matches 392; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MAGIPGLLFLFLCAGVQSPYSAWPKTPAYRLPVLPPOSTNLAKPDPGAEAKLE 60
 Db 1 MAGIPGLLFLFLCAGVQSPYSAWPKTPAYRLPVLPPOSTNLAKPDPGAEAKLE 60

Qy 61 VSSSCGPPOCHKGTPPLPKYKEAQYLSVETLYANGSRPEXOYIYLSSSGDAXXDSGS 120
 Db 61 VSSSCGPPOCHKGTPPLPKYKEAQYLSVETLYANGSRPEXOYIYLSSSGDAXXDSGS 120

Qy 121 SGSSRRKRQIYGDPSRSIFGKDFLNNYPFSTSVKLSTGCTLVEXHVTAAHC1HDG 180
 Db 121 SGSSRRKRQIYGDPSRSIFGKDFLNNYPFSTSVKLSTGCTLVEXHVTAAHC1HDG 180

Qy 181 KTYVKGTOKLRYGLPKFKDGGGRGANDSTSAMPQMKFQNTVRKTHVPKGWIKGNND 240
 Db 181 KTYVKGTOKLRYGLPKFKDGGGRGANDSTSAMPQMKFQNTVRKTHVPKGWIKGNND 240

Qy 241 IGMDYDALLEKKPHRKEMTIVGSPAKQPGGRHTFGCYDNDPFGNLYRFDYKDE 300
 Db 241 IGMDYDALLEKKPHRKEMTIVGSPAKQPGGRHTFGCYDNDPFGNLYRFDYKDE 300

Qy 301 TYDLYQQCDAQPGASGYGVYRMWKKRQQKWERKIGFESHOWDMNGSQDENYAVR 360
 Db 301 TYDLYQQCDAQPGASGYGVYRMWKKRQQKWERKIGFESHOWDMNGSQDENYAVR 360

Qy 361 ITPLKYAOICYWIKGNYLDCRGDTVFLPGSN 392
 Db 361 ITPLKYAOICYWIKGNYLDCRGDTVFLPGSN 392

RESULT 2
 US-09-072-384-15
 ; Sequence 18, Application US/09072384
 ; Patent No. 6153420
 ; GENERAL INFORMATION:
 ; APPLICANT: Sheppard, Paul O.
 ; TITLE OF INVENTION: SERINE PROTEASE POLYPEPTIDES
 ; TITLE OF INVENTION: AND MATERIALS AND METHODS FOR MAKING THEM
 ; NUMBER OF SEQUENCES: 18
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: ZymoGenetics Inc.
 ; STREET: 1201 Eastlake Avenue East
 ; CITY: Seattle
 ; STATE: WA
 ; ZIP: 98102
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: DOS
 ; SOFTWARE: FastSEQ for Windows Version 2.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/072, 384
 ; FILING DATE:
 ; PRIORITY APPLICATION DATA:
 ; APPLICATION NUMBER:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Parker, Gary E
 ; REGISTRATION NUMBER: 31, 648
 ; REFERENCE/DOCKET NUMBER: 97-16C1
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 206-442-6673
 ; TELEX:
 ; INFORMATION FOR SEQ ID NO: 15:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 392 amino acids

TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 FRAGMENT TYPE: internal
 FEATURE:
 NAME/KEY: Signal Sequence
 LOCATION: 1...19
 OTHER INFORMATION:
 US-09-072-384-15

Query Match 98.7%; Score 2085; DB 4; Length 392;
 Best Local Similarity 98.2%; Pred. No. 7.5e-21;
 Matches 385; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

Qy 1 MAGIPGLLFLFLCAGVQSPYSAWPKTPAYRLPVVLPOSTNLAKPDPGAEAKLE 60
 Db 1 MAGIPGLLFLFLCAGVQSPYSAWPKTPAYRLPVVLPOSTNLAKPDPGAEAKLE 60

Qy 61 VSSSCGPPOCHKGTPPLPKYKEAQYLSVETLYANGSRPEXOYIYLSSSGDAXXDSGS 120
 Db 61 VSSSCGPPOCHKGTPPLPKYKEAQYLSVETLYANGSRPEXOYIYLSSSGDQAQHRDGS 120

Qy 121 SGSSRRKRQIYGDPSRSIFGKDFLNNYPFSTSVKLSTGCTLVEXHVTAAHC1HDG 180
 Db 121 SGSSRRKRQIYGDPSRSIFGKDFLNNYPFSTSVKLSTGCTLVEXHVTAAHC1HDG 180

Qy 181 KTYVKGTOKLRYGLPKFKDGGGRGANDSTSAMPQMKFQNTVRKTHVPKGWIKGNND 240
 Db 181 KTYVKGTOKLRYGLPKFKDGGGRGANDSTSAMPQMKFQNTVRKTHVPKGWIKGNND 240

Qy 241 IGMDYDALLEKKPHRKEMTIVGSPAKQPGGRHTFGCYDNDPFGNLYRFDYKDE 300
 Db 241 IGMDYDALLEKKPHRKEMTIVGSPAKQPGGRHTFGCYDNDPFGNLYRFDYKDE 300

Qy 301 TYDLYQQCDAQPGASGYGVYRMWKKRQQKWERKIGFESHOWDMNGSQDENYAVR 360
 Db 301 TYDLYQQCDAQPGASGYGVYRMWKKRQQKWERKIGFESHOWDMNGSQDENYAVR 360

Qy 361 ITPLKYAOICYWIKGNYLDCRGDTVFLPGSN 392
 Db 361 ITPLKYAOICYWIKGNYLDCRGDTVFLPGSN 392

RESULT 3
 US-09-072-384-18
 ; Sequence 18, Application US/09072384
 ; Patent No. 6153420
 ; GENERAL INFORMATION:
 ; APPLICANT: Sheppard, Paul O.
 ; TITLE OF INVENTION: SERINE PROTEASE POLYPEPTIDES
 ; TITLE OF INVENTION: AND MATERIALS AND METHODS FOR MAKING THEM
 ; NUMBER OF SEQUENCES: 18
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: ZymoGenetics, Inc.
 ; STREET: 1201 Eastlake Avenue East
 ; CITY: Seattle
 ; STATE: WA
 ; ZIP: 98102
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: DOS
 ; SOFTWARE: FastSEQ for Windows Version 2.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/072, 384
 ; FILING DATE:
 ; PRIORITY APPLICATION DATA:
 ; APPLICATION NUMBER:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Parker, Gary E
 ; REGISTRATION NUMBER: 31, 648
 ; REFERENCE/DOCKET NUMBER: 97-16C1
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 206-442-6673
 ; TELEX:
 ; CLASSIFICATION:
 ; PRIORITY APPLICATION DATA:
 ; APPLICATION NUMBER:
 ; FILING DATE:

Result No.	Score	Query	Match	Length	DB ID	Description
1	2131	100.0	392	4	US-09-072-384-15	Sequence 15, Appl
2	2085	97.8	392	4	US-09-072-384-2	Sequence 2, Appl
3	2080	97.6	383	4	US-09-072-384-18	Sequence 18, Appl
4	177	8.3	222	1	US-08-090-048-1	Sequence 1, Appl
5	177	8.3	222	2	US-08-292-250-1	Sequence 1, Appl
6	177	8.3	222	2	US-07-927-631A-1	Sequence 1, Appl
7	122	5.7	256	3	US-08-906-769-89	Sequence 89, Appl
8	122	5.7	256	3	US-08-906-616-89	Sequence 89, Appl
9	122	5.7	256	4	US-08-817-795-89	Sequence 89, Appl
10	122	5.7	256	4	US-08-039-075A-89	Sequence 89, Appl
11	122	5.7	256	4	US-09-012-331-89	Sequence 89, Appl
12	122	5.7	256	4	US-09-032-215-32	Sequence 89, Appl
13	122	5.7	256	4	US-09-012-692-89	Sequence 89, Appl
14	114	5.3	253	6	5223425-8	Sequence 89, Appl
15	112	5.7	253	6	PCT-US5-1442A-89	Sequence 89, Appl
16	117	5.5	437	1	US-08-087-037-2	Sequence 2, Appl
17	115	5.4	241	4	US-08-944-83-59	Sequence 59, Appl
18	114	5.3	224	4	US-08-944-483-44	Sequence 4, Appl
19	114	5.3	253	6	5223425-5	Patent No. 5223425
20	113	5.3	250	6	5223425-4	Patent No. 5223425
21	113	5.3	223	1	US-08-278-091-13	Sequence 13, Appl
22	112.5	5.3	223	1	US-08-483-059-13	Sequence 13, Appl
23	112.5	5.3	223	1	US-08-483-059-13	Sequence 13, Appl
24	112.5	5.3	223	1	US-08-472-173-13	Sequence 13, Appl
25	112.5	5.3	223	2	US-08-487-167-13	Sequence 13, Appl
26	112.5	5.3	223	2	US-08-482-816-13	Sequence 13, Appl
27	112.5	5.3	223	2	US-08-296-149-13	Sequence 13, Appl

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2001, 16:15:33 ; Search time 35.78 Seconds
(without alignments)
246.542 Million cell updates/sec

Title: US-09-072-384-15
Perfect score: 2131
Sequence: 1 MAGIPGFLFLFLCAVGQ.....IKGNYLDRCGDTVFPFGSN 392

Scoring table: BLOSUM62
Gapoff 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA.*
1: /cgns_6/ptodata/2/laa/5A_COMB.pep:
2: /cgns_6/ptodata/2/laa/5B_COMB.pep:
3: /cgns_6/ptodata/2/laa/6A_COMB.pep:
4: /cgns_6/ptodata/2/laa/6B_COMB.pep:
5: /cgns_6/ptodata/2/laa/PCTUS_COMB.pep:
6: /cgns_6/ptodata/2/laa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Match	Length	DB ID	Description
1	100.0	392	4	US-09-072-384-15	Sequence 15, Appl	
2	97.8	392	4	US-09-072-384-2	Sequence 2, Appl	
3	97.6	383	4	US-09-072-384-18	Sequence 18, Appl	
4	177	222	1	US-08-090-048-1	Sequence 1, Appl	
5	177	222	2	US-08-292-250-1	Sequence 1, Appl	
6	177	222	2	US-07-927-631A-1	Sequence 1, Appl	
7	122	256	3	US-08-906-769-89	Sequence 89, Appl	
8	122	256	3	US-08-906-616-89	Sequence 89, Appl	
9	122	256	4	US-08-817-795-89	Sequence 89, Appl	
10	122	256	4	US-08-039-075A-89	Sequence 89, Appl	
11	122	256	4	US-09-012-331-89	Sequence 89, Appl	
12	122	256	4	US-09-032-215-32	Sequence 89, Appl	
13	122	256	4	US-09-012-692-89	Sequence 89, Appl	
14	114	253	6	5223425-8	Sequence 89, Appl	
15	112	253	6	PCT-US5-1442A-89	Sequence 89, Appl	
16	117	437	1	US-08-087-037-2	Sequence 2, Appl	
17	115	241	4	US-08-944-83-59	Sequence 59, Appl	
18	114	224	4	US-08-944-483-44	Sequence 4, Appl	
19	114	253	6	5223425-5	Patent No. 5223425	
20	113	250	6	5223425-4	Patent No. 5223425	
21	113	223	1	US-08-278-091-13	Sequence 13, Appl	
22	112.5	223	1	US-08-483-059-13	Sequence 13, Appl	
23	112.5	223	1	US-08-483-059-13	Sequence 13, Appl	
24	112.5	223	1	US-08-472-173-13	Sequence 13, Appl	
25	112.5	223	2	US-08-487-167-13	Sequence 13, Appl	
26	112.5	223	2	US-08-482-816-13	Sequence 13, Appl	
27	112.5	223	2	US-08-296-149-13	Sequence 13, Appl	

RESULT 1
US-09-072-384-15
; Sequence 15, Application US/09072384
; Patent No. 6153420

GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; TITLE OF INVENTION: SERINE PROTEASE POLYPEPTIDES
; TITLE OF INVENTION: AND MATERIALS AND METHODS FOR MAKING THEM
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESS: ZymoGenetics, Inc.
; STREET: 1201 Eastlake Avenue East
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98102

COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072_384
; FILING DATE:
; CLASSIFICATION:
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, Gary E.
; REGISTRATION NUMBER: 31,648
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-442-6673
; TELEFAX: 206-442-6678
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 392 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE: internal
; NAME/KEY: Signal Sequence
; LOCATION: 1...19
; OTHER INFORMATION:
; US-09-072-384-15

Query Match 100.08; Score 2131; DB 4; Length 392;
 Best Local Similarity 100.08; Pred. No. 1..1e-27;
 Matches 392; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MAGIPGLLFLFLLCAGVQSPYSAWPKTPWAPRLPVVLPOSTLNALKPDPGAEAKLE 60
 Db 1 MAGIPGLLFLFLLCAGVQSPYSAWPKTPWAPRLPVVLPOSTLNALKPDPGAEAKLE 60

Qy 61 VSSSGPQCHKGTPPLPIYEAAQYLSEVETLYANGSRPETOQYIYLSSSGDGAQHRDSGS 120
 Db 61 VSSSGPQCHKGTPPLPIYEAAQYLSEVETLYANGSRPETOQYIYLSSSGDGAQHRDSGS 120

Qy 121 SGKSRRKRQIYGDKDSRISIFGIDFLNYPFSTSVKLSTGCTGTLVAEKVLTAAHC1HDG 180
 Db 121 SGKSRRKRQIYGDKDSRISIFGIDFLNYPFSTSVKLSTGCTGTLVAEKVLTAAHC1HDG 180

Qy 181 KTYVKGTQKLRVGLPKFKDGGGRANDSTSAMPQMKFQWTRVKRTHVPKGWIKGNND 240
 Db 181 KTYVKGTQKLRVGLPKFKDGGGRANDSTSAMPQMKFQWTRVKRTHVPKGWIKGNND 240

Qy 241 IGMDDYALLEKKPHRKEMITIGVSPAKOLPGGRITHFSCYDNDPQGNLYRFCDVYKDE 300
 Db 241 IGMDDYALLEKKPHRKEMITIGVSPAKOLPGGRITHFSCYDNDPQGNLYRFCDVYKDE 300

Qy 301 TYDLIYQCDAGPGASGSQGYVRMVKRQQKWERKLTIGFSQHWDMMNGSPQDFNVAVR 360
 Db 301 TYDLIYQCDAGPGASGSQGYVRMVKRQQKWERKLTIGFSQHWDMMNGSPQDFNVAVR 360

Qy 361 ITPLKYAQICYWKGNYLDCREGDTVPPGSN 392
 Db 361 ITPLKYAQICYWKGNYLDCREGDTVPPGSN 392

RESULT 2
 US-09-072-384-2
 ; Sequence 2, Application US/09072384
 ; Patent No. 6153420
 ; GENERAL INFORMATION:
 ; APPLICANT: Sheppard, Paul O.
 ; TITLE OF INVENTION: SERINE PROTEASE POLYPEPTIDES
 ; TITLE OF INVENTION: AND MATERIALS AND METHODS FOR MAKING THEM
 ; NUMBER OF SEQUENCES: 18
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: ZymoGenetics, Inc.
 ; STREET: 1201 Eastlake Avenue East
 ; CITY: Seattle
 ; STATE: WA
 ; ZIP: 98102
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: DOS
 ; SOFTWARE: FastSEQ for Windows Version 2.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/072, 384
 ; FILING DATE:
 ; PRIORITY APPLICATION DATA:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Parker, Gary E
 ; REGISTRATION NUMBER: 31, 648
 ; REFERENCE/DOCKET NUMBER: 97-16C1
 ; TELEPHONE: 206-442-6673
 ; TELEX: 206-442-6678
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 392 amino acids

; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; FRAGMENT TYPE: internal
 ; FEATURE:
 ; NAME/KEY: Signal Sequence
 ; LOCATION: 1...19
 ; OTHER INFORMATION:
 ; US-09-072-384-2

Query Match 97.88; Score 2085; DB 4;
 Best Local Similarity 98.28;
 Matches 385; Conservative 98.28;
 ; Mismatches 6; Indels 0; Gaps 0;

Qy 1 MAGIPGLLFLFLLCAGVQSPYSAWPKTPWAPRLPVVLPOSTLNALKPDPGAEAKLE 60
 Db 1 MAGIPGLLFLFLLCAGVQSPYSAWPKTPWAPRLPVVLPOSTLNALKPDPGAEAKLE 60

Qy 61 VSSSGPQCHKGTPPLPIYEAAQYLSEVETLYANGSRPETOQYIYLSSSGDGAQHRDSGS 120
 Db 61 VSSSGPQCHKGTPPLPIYEAAQYLSEVETLYANGSRPETOQYIYLSSSGDGAQHRDSGS 120

Qy 121 SGKSRRKRQIYGDKDSRISIFGIDFLNYPFSTSVKLSTGCTGTLVAEKVLTAAHC1HDG 180
 Db 121 SGKSRRKRQIYGDKDSRISIFGIDFLNYPFSTSVKLSTGCTGTLVAEKVLTAAHC1HDG 180

Qy 181 KTYVKGTQKLRVGLPKFKDGGGRANDSTSAMPQMKFQWTRVKRTHVPKGWIKGNND 240
 Db 181 KTYVKGTQKLRVGLPKFKDGGGRANDSTSAMPQMKFQWTRVKRTHVPKGWIKGNND 240

Qy 241 IGMDDYALLEKKPHRKEMITIGVSPAKOLPGGRITHFSCYDNDPQGNLYRFCDVYKDE 300
 Db 241 IGMDDYALLEKKPHRKEMITIGVSPAKOLPGGRITHFSCYDNDPQGNLYRFCDVYKDE 300

Qy 301 TYDLIYQCDAGPGASGSQGYVRMVKRQQKWERKLTIGFSQHWDMMNGSPQDFNVAVR 360
 Db 301 TYDLIYQCDAGPGASGSQGYVRMVKRQQKWERKLTIGFSQHWDMMNGSPQDFNVAVR 360

Qy 361 ITPLKYAQICYWKGNYLDCREGDTVPPGSN 392
 Db 361 ITPLKYAQICYWKGNYLDCREGDTVPPGSN 392

RESULT 3
 US-09-072-384-18
 ; Sequence 18, Application US/09072384
 ; Patent No. 6153420
 ; GENERAL INFORMATION:
 ; APPLICANT: Sheppard, Paul O.
 ; TITLE OF INVENTION: SERINE PROTEASE POLYPEPTIDES
 ; NUMBER OF SEQUENCES: 18
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: ZymoGenetics, Inc.
 ; STREET: 1201 Eastlake Avenue East
 ; CITY: Seattle
 ; STATE: WA
 ; COUNTRY: USA
 ; ZIP: 98102
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: DOS
 ; SOFTWARE: FastSEQ for Windows Version 2.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/072, 384
 ; FILING DATE:
 ; PRIORITY APPLICATION DATA:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Parker, Gary E
 ; REGISTRATION NUMBER: 31, 648
 ; REFERENCE/DOCKET NUMBER: 97-16C1
 ; TELEPHONE: 206-442-6673
 ; TELEX: 206-442-6678
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/072, 384
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER:
 ; FILING DATE:

need 16-19

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

Om protein - protein search, using sw model

Run on: December 2, 2001, 16:14:40 ; Search time 64.47 Seconds

(without alignments)
450 391 Million cell updates/sec

Title: US-09-072-384-2

Perfect score: 2112

Sequence: 1 MAGIPGLLFFLLCAVGGQ IKGNYLDCREGDTVFLPGSN 392

Scoring table: BLOSUM62
GapP 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters:

522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_1101;*

1: /SIDS2/gcadata/geneseq/geneseqp/AA1980.DAT;*
2: /SIDS2/gcadata/geneseq/geneseqp/AA1981.DAT;*
3: /SIDS2/gcadata/geneseq/geneseqp/AA1982.DAT;*
4: /SIDS2/gcadata/geneseq/geneseqp/AA1984.DAT;*
5: /SIDS2/gcadata/geneseq/geneseqp/AA1985.DAT;*
6: /SIDS2/gcadata/geneseq/geneseqp/AA1986.DAT;*
7: /SIDS2/gcadata/geneseq/geneseqp/AA1987.DAT;*
8: /SIDS2/gcadata/geneseq/geneseqp/AA1988.DAT;*
9: /SIDS2/gcadata/geneseq/geneseqp/AA1989.DAT;*
10: /SIDS2/gcadata/geneseq/geneseqp/AA1990.DAT;*
11: /SIDS2/gcadata/geneseq/geneseqp/AA1991.DAT;*
12: /SIDS2/gcadata/geneseq/geneseqp/AA1992.DAT;*
13: /SIDS2/gcadata/geneseq/geneseqp/AA1993.DAT;*
14: /SIDS2/gcadata/geneseq/geneseqp/AA1994.DAT;*
15: /SIDS2/gcadata/geneseq/geneseqp/AA1995.DAT;*
16: /SIDS2/gcadata/geneseq/geneseqp/AA1996.DAT;*
17: /SIDS2/gcadata/geneseq/geneseqp/AA1997.DAT;*
18: /SIDS2/gcadata/geneseq/geneseqp/AA1998.DAT;*
19: /SIDS2/gcadata/geneseq/geneseqp/AA1999.DAT;*
20: /SIDS2/gcadata/geneseq/geneseqp/AA2000.DAT;*
21: /SIDS2/gcadata/geneseq/geneseqp/AA2001.DAT;*
22: /SIDS2/gcadata/geneseq/geneseqp/AA2002.DAT;*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

ALIGNMENTS

RESULT 1

AAB48972 standard; Protein; 392 AA.

ID XX

AC AAB48972;

XX

DT 27-MAR-2001 (first entry)

XX

DE Human Zsig13 variant #1, SEQ ID NO:2.

XX

KW Human Zsig13; serine protease; chromosome 11q22.1; elastase homologue; glutamyl endopeptidase homologue; factor X homologue; trypsin homologue;

KW trypsinogen homologue; mast cell protease homologue;

KW collagenase homologue; protein degradation homologue;

KW alcohol production; laundry detergent component.

OS Homo sapiens.

XX

PN US615420-A.

XX

PD 28-NOV-2000.

XX

PF 04-MAY-1998;

XX

PR 24-APR-1997;

XX

PR 17-APR-1998;

XX

PA (ZYMO) ZYMOGENETICS INC.

XX

PI Sheppard PO; - inactivator

XX

DR WPI: 2001-060090/07.

XX N-PSDB; AAC91782.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2104	99.6	392	22 AAB48972	Human Zsig13 varia
2	2085	98.7	392	22 AAB48973	Human Zsig13 varia
3	2044	96.8	383	20 AAY08657	Human transmembran
4	2044	96.8	383	20 AAY08650	W09927094 Seq ID 1
5	2044	96.8	383	20 AAY13390	Amino acid sequenc
6	2044	96.8	383	21 AAB25592	Protein encoded by
7	2044	96.8	383	21 AAB25618	Human TANGO 186 pr
8	2044	96.8	383	21 AAY8727	Human signal Pepti
9	2044	96.8	383	21 AAY5327	A_bone mario secr
10	2044	96.8	383	22 AAB80258	Human PRO307 prote
11	2044	96.8	383	22	

Scoring table:	BLOSUM62	ALIGNMENTS			
Searched:	100059 seqs, 36664827 residues	RESULT	1		
Total number of hits satisfying chosen parameters:	100059	ID	GSEP_BACLI	STANDARD;	PRT; 316 AA.
Minimum DB seq length: 0		ID	GSEP_BACLI	STANDARD;	PRT; 316 AA.
Maximum DB seq length: 2000000000		AC	P80057;		
Post-processing: Minimum Match 0%		DT	01-MAY-1992 (Rel. 22, Created)		
Post-processing: Maximum Match 100%		DT	01-OCT-1993 (Rel. 27, Last sequence update)		
Post-processing: Listing first 45 summaries		DT	20-AUG-2001 (Rel. 40, Last annotation update)		
Database :	SwissProt_39.*	DE	GLUTANYL ENDopeptidase PRECURSOR (EC 3.4.21.19) (GLUTAMATE SPECIFIC ENDopeptidase) (GSE).		
Scoring table:	BLOSUM62	GN	glutamyl acid-specific protease from <i>Bacillus licheniformis</i> .		
Gappp:	10.0 , capext 0.5	OS	<i>Bacillus licheniformis</i> .		
Searched:	100059 seqs, 36664827 residues	OC	<i>Bacteria</i> ; <i>Firmicutes</i> ; <i>Bacillus/Clostridium</i> group;		
Total number of hits satisfying chosen parameters:	100059	OC	<i>Bacillus</i> ; <i>Staphylococcus</i> group; <i>Bacillus</i> .		
Minimum DB seq length: 0		RN	[1]		
Maximum DB seq length: 2000000000		RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.		
Post-processing: Minimum Match 0%		RC	STRAIN=TCC 14580; PubMed=1429718;		
Post-processing: Maximum Match 100%		RX	MEDLINE=93054737;		
Post-processing: Listing first 45 summaries		RA	Kakudo S., Kikuchi N., Kitadokoro K., Fujiwara T., Nakamura E., Teraoka H., Yoshida N.;		
Database :	SwissProt_39.*	RA	Okamoto H., Shin M., Tamaki M., Tsuzuki H., Yoshida N.;		
Scoring table:	BLOSUM62	RT	"Purification, characterization, cloning, and expression of a		
Gappp:	10.0 , capext 0.5	RA	glutamyl acid-specific protease from <i>Bacillus licheniformis</i> ATCC 14580."		
Searched:	100059 seqs, 36664827 residues	RT	RT		
Total number of hits satisfying chosen parameters:	100059	RL	J. Biol. Chem. 267:23782-23788(1992).		
Minimum DB seq length: 0		RN	[2]		
Maximum DB seq length: 2000000000		RP	SEQUENCE 95-316.		
Post-processing: Minimum Match 0%		RX	MEDLINE=92155199; PubMed=1346764;		
Post-processing: Maximum Match 100%		RA	Svendsen I., Breedam K.;		
Post-processing: Listing first 45 summaries		RT	"Isolation and amino acid sequence of a glutamic acid specific endopeptidase from <i>Bacillus licheniformis</i> .";		
Database :	SwissProt_39.*	RL	Eur. J. Biochem. 204:165-171(1992).		
Scoring table:	BLOSUM62	CC	-I- FUNCTION: SPECIFIC FOR HYDROLYSIS OF PEPTIDES BONDS ON THE CARBOXYL SIDE OF ACIDIC AMINO ACID RESIDUES, WITH A STRONG PREFERENCE FOR GLU.		
Gappp:	10.0 , capext 0.5	CC	-I- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE: ASP-1-XAA, GLU-1-XAA.		
Searched:	100059 seqs, 36664827 residues	CC	-I- SUBCELLULAR LOCATION: SECRETED.		
Total number of hits satisfying chosen parameters:	100059	CC	-I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2B; ALSO KNOWN AS THE V8 FAMILY.		
Minimum DB seq length: 0		CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See: http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).		
Maximum DB seq length: 2000000000		CC	CC		
Post-processing: Minimum Match 0%		CC	CC		
Post-processing: Maximum Match 100%		CC	CC		
Post-processing: Listing first 45 summaries		CC	CC		
Database :	SwissProt_39.*	CC	CC		
Scoring table:	BLOSUM62	CC	CC		
Gappp:	10.0 , capext 0.5	CC	CC		
Searched:	100059 seqs, 36664827 residues	CC	CC		
Total number of hits satisfying chosen parameters:	100059	CC	CC		
Minimum DB seq length: 0		CC	CC		
Maximum DB seq length: 2000000000		CC	CC		
Post-processing: Minimum Match 0%		CC	CC		
Post-processing: Maximum Match 100%		CC	CC		
Post-processing: Listing first 45 summaries		CC	CC		
Database :	SwissProt_39.*	CC	CC		
Scoring table:	BLOSUM62	CC	CC		
Gappp:	10.0 , capext 0.5	CC	CC		
Searched:	100059 seqs, 36664827 residues	CC	CC		
Total number of hits satisfying chosen parameters:	100059	CC	CC		
Minimum DB seq length: 0		CC	CC		
Maximum DB seq length: 2000000000		CC	CC		
Post-processing: Minimum Match 0%		CC	CC		
Post-processing: Maximum Match 100%		CC	CC		
Post-processing: Listing first 45 summaries		CC	CC		
Database :	SwissProt_39.*	CC	CC		
Scoring table:	BLOSUM62	CC	CC		
Gappp:	10.0 , capext 0.5	CC	CC		
Searched:	100059 seqs, 36664827 residues	CC	CC		
Total number of hits satisfying chosen parameters:	100059	CC	CC		
Minimum DB seq length: 0		CC	CC		
Maximum DB seq length: 2000000000		CC	CC		
Post-processing: Minimum Match 0%		CC	CC		
Post-processing: Maximum Match 100%		CC	CC		
Post-processing: Listing first 45 summaries		CC	CC		
Database :	SwissProt_39.*	CC	CC		
Scoring table:	BLOSUM62	CC	CC		
Gappp:	10.0 , capext 0.5	CC	CC		
Searched:	100059 seqs, 36664827 residues	CC	CC		
Total number of hits satisfying chosen parameters:	100059	CC	CC		
Minimum DB seq length: 0		CC	CC		
Maximum DB seq length: 2000000000		CC	CC		
Post-processing: Minimum Match 0%		CC	CC		
Post-processing: Maximum Match 100%		CC	CC		
Post-processing: Listing first 45 summaries		CC	CC		
Database :	SwissProt_39.*	CC	CC		
Scoring table:	BLOSUM62	CC	CC		
Gappp:	10.0 , capext 0.5	CC	CC		
Searched:	100059 seqs, 36664827 residues	CC	CC		
Total number of hits satisfying chosen parameters:	100059	CC	CC		
Minimum DB seq length: 0		CC	CC		
Maximum DB seq length: 2000000000		CC	CC		
Post-processing: Minimum Match 0%		CC	CC		
Post-processing: Maximum Match 100%		CC	CC		
Post-processing: Listing first 45 summaries		CC	CC		
Database :	SwissProt_39.*	CC	CC		
Scoring table:	BLOSUM62	CC	CC		
Gappp:	10.0 , capext 0.5	CC	CC		
Searched:	100059 seqs, 36664827 residues	CC	CC		
Total number of hits satisfying chosen parameters:	100059	CC	CC		
Minimum DB seq length: 0		CC	CC		
Maximum DB seq length: 2000000000		CC	CC		
Post-processing: Minimum Match 0%		CC	CC		
Post-processing: Maximum Match 100%		CC	CC		
Post-processing: Listing first 45 summaries		CC	CC		
Database :	SwissProt_39.*	CC	CC		
Scoring table:	BLOSUM62	CC	CC		
Gappp:	10.0 , capext 0.5	CC	CC		
Searched:	100059 seqs, 36664827 residues	CC	CC		
Total number of hits satisfying chosen parameters:	100059	CC	CC		
Minimum DB seq length: 0		CC	CC		
Maximum DB seq length: 2000000000		CC	CC		
Post-processing: Minimum Match 0%		CC	CC		
Post-processing: Maximum Match 100%		CC	CC		
Post-processing: Listing first 45 summaries		CC	CC		
Database :	SwissProt_39.*	CC	CC		
Scoring table:	BLOSUM62	CC	CC		
Gappp:	10.0 , capext 0.5	CC	CC		
Searched:	100059 seqs, 36664827 residues	CC	CC		
Total number of hits satisfying chosen parameters:	100059	CC	CC		
Minimum DB seq length: 0		CC	CC		
Maximum DB seq length: 2000000000		CC	CC		
Post-processing: Minimum Match 0%		CC	CC		
Post-processing: Maximum Match 100%		CC	CC		
Post-processing: Listing first 45 summaries		CC	CC		
Database :	SwissProt_39.*	CC	CC		
Scoring table:	BLOSUM62	CC	CC		
Gappp:	10.0 , capext 0.5	CC	CC		
Searched:	100059 seqs, 36664827 residues	CC	CC		
Total number of hits satisfying chosen parameters:	100059	CC	CC		
Minimum DB seq length: 0		CC	CC		
Maximum DB seq length: 2000000000		CC	CC		
Post-processing: Minimum Match 0%		CC	CC		
Post-processing: Maximum Match 100%		CC	CC		
Post-processing: Listing first 45 summaries		CC	CC		
Database :	SwissProt_39.*	CC	CC		
Scoring table:	BLOSUM62	CC	CC		
Gappp:	10.0 , capext 0.5	CC	CC		
Searched:	100059 seqs, 36664827 residues	CC	CC		
Total number of hits satisfying chosen parameters:	100059	CC	CC		
Minimum DB seq length: 0		CC	CC		
Maximum DB seq length: 2000000000		CC	CC		
Post-processing: Minimum Match 0%		CC	CC		
Post-processing: Maximum Match 100%		CC	CC		
Post-processing: Listing first 45 summaries		CC	CC		
Database :	SwissProt_39.*	CC	CC		
Scoring table:	BLOSUM62	CC	CC		
Gappp:	10.0 , capext 0.5	CC	CC		
Searched:	100059 seqs, 36664827 residues	CC	CC		
Total number of hits satisfying chosen parameters:	100059	CC	CC		
Minimum DB seq length: 0		CC	CC		
Maximum DB seq length: 2000000000		CC	CC		
Post-processing: Minimum Match 0%		CC	CC		
Post-processing: Maximum Match 100%		CC	CC		
Post-processing: Listing first 45 summaries		CC	CC		
Database :	SwissProt_39.*	CC	CC		
Scoring table:	BLOSUM62	CC	CC		
Gappp:	10.0 , capext 0.5	CC	CC		
Searched:	100059 seqs, 36664827 residues	CC	CC		
Total number of hits satisfying chosen parameters:	100059	CC	CC		
Minimum DB seq length: 0		CC	CC		
Maximum DB seq length: 2000000000		CC	CC		
Post-processing: Minimum Match 0%		CC	CC		
Post-processing: Maximum Match 100%		CC	CC		
Post-processing: Listing first 45 summaries		CC	CC		
Database :	SwissProt_39.*	CC	CC		
Scoring table:	BLOSUM62	CC	CC		
Gappp:	10.0 , capext 0.5	CC	CC		
Searched:	100059 seqs, 36664827 residues	CC	CC		
Total number of hits satisfying chosen parameters:	100059	CC	CC		
Minimum DB seq length: 0		CC	CC		
Maximum DB seq length: 2000000000		CC	CC		
Post-processing: Minimum Match 0%		CC	CC		
Post-processing: Maximum Match 100%		CC	CC		
Post-processing: Listing first 45 summaries		CC	CC		
Database :	SwissProt_39.*	CC	CC		
Scoring table:	BLOSUM62	CC	CC		
Gappp:	10.0 , capext 0.5	CC	CC		
Searched:	100059 seqs, 36664827 residues	CC	CC		
Total number of hits satisfying chosen parameters:	100059	CC	CC		
Minimum DB seq length: 0		CC	CC		
Maximum DB seq length: 2000000000		CC	CC		
Post-processing: Minimum Match 0%		CC	CC		
Post-processing: Maximum Match 100%		CC	CC		
Post-processing: Listing first 45 summaries		CC	CC		
Database :	SwissProt_39.*	CC	CC		
Scoring table:	BLOSUM62	CC	CC		
Gappp:	10.0 , capext 0.5	CC	CC		
Searched:	100059 seqs, 36664827 residues	CC	CC		
Total number of hits satisfying chosen parameters:	100059	CC	CC		
Minimum DB seq length: 0		CC	CC		
Maximum DB seq length: 2000000000		CC	CC		
Post-processing: Minimum Match 0%		CC	CC		
Post-processing: Maximum Match 100%		CC	CC		
Post-processing: Listing first 45 summaries		CC	CC		
Database :	SwissProt_39.*	CC	CC		
Scoring table:	BLOSUM62	CC	CC		
Gappp:	10.0 , capext 0.5	CC	CC		
Searched:	100059 seqs, 36664827 residues	CC	CC		
Total number of hits satisfying chosen parameters:	100059	CC	CC		
Minimum DB seq length: 0		CC	CC		
Maximum DB seq length: 2000000000		CC	CC		
Post-processing: Minimum Match 0%		CC	CC		
Post-processing: Maximum Match 100%		CC	CC		
Post-processing: Listing first 45 summaries		CC	CC		
Database :	SwissProt_39.*	CC	CC		
Scoring table:	BLOSUM62	CC	CC		
Gappp:	10.0 , capext 0.5	CC	CC		
Searched:	100059 seqs, 36664827 residues	CC	CC		
Total number of hits satisfying chosen parameters:	100059	CC	CC		
Minimum DB seq length: 0		CC	CC		
Maximum DB seq length: 2000000000		CC	CC		
Post-processing: Minimum Match 0%		CC	CC		
Post-processing: Maximum Match 100%		CC	CC		
Post-processing: Listing first 45 summaries		CC	CC		
Database :	SwissProt_39.*	CC	CC		
Scoring table:	BLOSUM62	CC	CC		
Gappp:	10.0 , capext 0.5	CC	CC		
Searched:	100059 seqs, 36664827 residues	CC	CC		
Total number of hits satisfying chosen parameters:	100059	CC	CC		
Minimum DB seq length: 0		CC	CC		
Maximum DB seq length: 2000000000		CC	CC		
Post-processing: Minimum Match 0%		CC	CC		
Post-processing: Maximum Match 100%		CC	CC		
Post-processing: Listing first 45 summaries		CC	CC		
Database :	SwissProt_39.*	CC	CC		
Scoring table:	BLOSUM62	CC	CC		
Gappp:	10.0 , capext 0.5	CC	CC		
Searched:	100059 seqs, 36664827 residues	CC	CC		
Total number of hits satisfying chosen parameters:	100059	CC	CC		
Minimum DB seq length: 0		CC	CC		
Maximum DB seq length: 2000000000		CC	CC		
Post-processing: Minimum Match 0%		CC	CC		
Post-processing: Maximum Match 100%		CC	CC		
Post-processing: Listing first 45 summaries		CC	CC		
Database :	SwissProt_39.*	CC	CC		
Scoring table:	BLOSUM62	CC	CC		
Gappp:	10.0 , capext 0.5	CC	CC		
Searched:	100059 seqs, 36664827 residues	CC	CC		
Total number of hits satisfying chosen parameters:	100059	CC	CC	</td	

DR	PROSTEIN; PS00672;	V8_HIS; 1.	RX MEDLINE-8910019; PubMed=3145905;
KW	Hydrolase; Serine protease; Signal	V8_SER; 1.	RA Smith H., de Jong A., Bron S., Venema G.; RT "Characterization of a partial sequence-coding region selected from the Bacillus subtilis chromosome."; RT Gene 70:351-361(1988).
FT	SIGNAL	?	RT
FT	PROPEP	?	RT
FT	CHAIN	95	RNA
FT	ACT_SITE	141	SEQUENCE OF 1-10 FROM N.A.
FT	ACT_SITE	261	RP STRAIN=168;
FT	DISULFID	126	RC SUBCELLULAR LOCATION: SECRETED.
FT	DISULFID	275	CC -1 SIMILARITY: BELOWS TO PEPTIDE FAMILY S2B; ALSO KNOWN AS THE V8 FAMILY.
FT	SEQUENCE	316 AA;	CC -1 CAUTION: CALLED "METALLOPROTEASE", BUT CLEARLY BELONGS TO THE S2 FAMILY OF SERINE PROTEASES.
Db	SSCGPOCHKGTPPL--PTYKEAKOYLSYETLYANGSRTEXQGYIYLSSGDGAXXRDSG 119	CC	CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation of the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no modified and this statement is not removed. Usage by commercial entities requires a license agreement (See http://www.ebi.ac.uk/announce or send an email to license@ebi.ac.uk).
Db	28 AQAAASPPH--TPVSSDPSYK AETSYVYDP ---NIKSDOGLYLSKAFTGTTG ---KVNE 76	CC	CC DR L10505; AAA22604.1; .
Db	63 SS CGKSRRK-----RQIYGDSRSFSIFGKDFLNYPFESTSVKLST---GCTGTVAE 167	CC	DR EMBL; AB00624; BAA3312.1; .
Db	77 TEKAEKSKPAKAPSIKSIVGKDDTRVTN---TAYPAPRATHISSSSGCTCWTMIGP 133	CC	DR EMBL; M22916; AAA22832.1; .
Db	120 XHVLTAAHC1HDGKT YVKGTQKLRYVGFLPKFKDGGRGANDSTSAMPOMKFQWIRVKR 226	CC	DR SMART; SM00020; TRYPSIN; .
Db	134 KTVATAGHC1YDTSSGSFAGTAVSPG-----RNCTS----YPYGSVKSTR 175	CC	DR PIR; A51122; A35122.
Db	168 THVPKGWIKGNANDIGMDYVALLEDKKPHKRKENKIGVSPAKOLPGGRTHFSGYDNDR 286	CC	DR HSSP; P00756; 1SGF.
Db	176 YFIPSGWRSGNNT----YDGAELSEPIGNITVGYFCSYTSSILVGTIVTISGPGDK 230	CC	DR MEROPS; S01_1272; .
Db	227 PGNLVYRFCD--VKDETDIYQQCDAQPGASGYGVYKRWKQROQQWERKII---GIF 340	CC	DR SubtilList; BG10690; mpr.
Db	231 TAGTQWHSGSPKIASETYKLOYAM-DTYGGOSGSPVFEQSSSRTRNGSPSLAVHTNGVY 289	CC	DR InterPro; IPR00126; Ser_proteas_V8.
Db	341 SGHQWYDMNGSPQDENVAVRTPLKYAQICYW 372	CC	DR PFam; PF00089; trypsin; .
Db	290 GG-----SSYNRGTRITKEVEDNLTNW 311	CC	DR PRINTS; PRO0839; V8PROTEASE.
Db	290 GG-----SSYNRGTRITKEVEDNLTNW 311	CC	DR SMART; SM00020; TRYPSIN; .
Db	290 GG-----SSYNRGTRITKEVEDNLTNW 311	CC	DR PROSITE; PS00672; V8_HIS; .
Db	290 GG-----SSYNRGTRITKEVEDNLTNW 311	CC	DR PROSITE; PS00673; V8_SER; .
Db	290 GG-----SSYNRGTRITKEVEDNLTNW 311	CC	DR KW Hydrolase; Serine Protease; Signal1; Zymogen; Complete proteome.
Db	290 GG-----SSYNRGTRITKEVEDNLTNW 311	CC	DR SIGNAL 1 34 POTENTIAL.
Db	290 GG-----SSYNRGTRITKEVEDNLTNW 311	CC	DR FT PROPEP 35 93 EXTRACELLULAR METALLOPROTEASE.
Db	290 GG-----SSYNRGTRITKEVEDNLTNW 311	CC	DR FT CHAIN 94 313 CHARGE RELAY SYSTEM (BY SIMILARITY).
Db	290 GG-----SSYNRGTRITKEVEDNLTNW 311	CC	DR FT ACT_SITE 146 146 CHARGE RELAY SYSTEM (BY SIMILARITY).
Db	290 GG-----SSYNRGTRITKEVEDNLTNW 311	CC	DR FT ACT_SITE 267 267 BY SIMILARITY.
Db	290 GG-----SSYNRGTRITKEVEDNLTNW 311	CC	DR FT DISULFID 131 147 OVSAPYEG > PLESTAOA (IN REF. 3).
Db	290 GG-----SSYNRGTRITKEVEDNLTNW 311	CC	DR FT CONFLICT 61 68 SQ SEQUENCE 313 AA; 33842 MW; D417886E8D52AE94 CRC64;
Db	290 GG-----SSYNRGTRITKEVEDNLTNW 311	CC	DR Query Match 6.28; Score 130.5 ; DB 1; Length 313; Best Local Similarity 22.6%; Pred. No. 0; 0/0038; Gaps 19; Matches 87; Conservative 38; Mismatches 139; Indels 121; Gaps 19
Qy	01-FEB-1995 (Rel. 31, Created)	Qy 28 WKPTWPAYRLPVLPQSTLNLAKPDPFGMAKLEVS---SSCGPQCHKGTPLPYKEAK 82	
Qy	01-FEB-1995 (Rel. 31, Last sequence update)	Db 7 ERKQWFAYLTVCLALA---AAVSEFSPVAKAENPQTSVNTGKEA-DATKNOTSADQ 61	
Qy	20-AUG-2001 (Rel. 40, Last annotation update)	Qy 83 QLYSYE-----TLYANGSRTEXQSYTILSSSGDGAAXRDSGSSGRSRRKQRIYGD 135	
Qy	DETRACELLULAR METALLOPROTEASE PRECURSOR (EC 3.4.21.-).	Db 62 VSAPYEGTKTSKSLY-GGQELEKNQTLOPS---.SIIGTDE 100	
GN	MPR.	Qy 136 RFSIFGKDPELLNPFSVSKLST-----GCTGTIVAEKHLTAACHTH-----178	
OS	Bacillus subtilis	Qy 136 RFSIFGKDPELLNPFSVSKLST-----GCTGTIVAEKHLTAACHTH-----178	
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;	Db 1.01 RTRI---SSTISFSPYRATQLKTPNTSTSCTGTFIVNPVTTVAGCHSYQSQDHWAS 157	
OC	Bacillus/Staphylococcus group; Bacillus.	Db 1.01 RTRI---SSTISFSPYRATQLKTPNTSTSCTGTFIVNPVTTVAGCHSYQSQDHWAS 157	
NCBI_TAXID=1423;		Db 1.01 RTRI---SSTISFSPYRATQLKTPNTSTSCTGTFIVNPVTTVAGCHSYQSQDHWAS 157	
RN	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.	Db 1.01 RTRI---SSTISFSPYRATQLKTPNTSTSCTGTFIVNPVTTVAGCHSYQSQDHWAS 157	
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.	Db 1.01 RTRI---SSTISFSPYRATQLKTPNTSTSCTGTFIVNPVTTVAGCHSYQSQDHWAS 157	
STRAIN=GP241;		Db 1.01 RTRI---SSTISFSPYRATQLKTPNTSTSCTGTFIVNPVTTVAGCHSYQSQDHWAS 157	
RX	MEDLINE-90130256; PubMed=2105291;	Db 1.01 RTRI---SSTISFSPYRATQLKTPNTSTSCTGTFIVNPVTTVAGCHSYQSQDHWAS 157	
RA	Sloma A., Rudolph C.F., Rufo G.A. Jr., Sullivan B.J., Theriault K.A., Aliy D., Pero J., "Gene encoding a novel extracellular metalloprotease in Bacillus subtilis"; J. Bacteriol. 172:1024-1029(1990).	Db 1.01 RTRI---SSTISFSPYRATQLKTPNTSTSCTGTFIVNPVTTVAGCHSYQSQDHWAS 157	
RT	[2]	Db 1.01 RTRI---SSTISFSPYRATQLKTPNTSTSCTGTFIVNPVTTVAGCHSYQSQDHWAS 157	
RT	SEQUENCE FROM N.A.	Db 1.01 RTRI---SSTISFSPYRATQLKTPNTSTSCTGTFIVNPVTTVAGCHSYQSQDHWAS 157	
RT	STRAIN=168;	Db 1.01 RTRI---SSTISFSPYRATQLKTPNTSTSCTGTFIVNPVTTVAGCHSYQSQDHWAS 157	
RA	Haga K., Liu H., Yasumoto K., takahashi H., Yoshioka H.; "Sequence analysis of the 70kb region between 17 and 23 degree of the Bacillus subtilis chromosome.", Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.	Db 1.01 RTRI---SSTISFSPYRATQLKTPNTSTSCTGTFIVNPVTTVAGCHSYQSQDHWAS 157	
RT	[3]	Db 1.01 RTRI---SSTISFSPYRATQLKTPNTSTSCTGTFIVNPVTTVAGCHSYQSQDHWAS 157	
RP	SEQUENCE OF 1-58 FROM N.A.	Db 1.01 RTRI---SSTISFSPYRATQLKTPNTSTSCTGTFIVNPVTTVAGCHSYQSQDHWAS 157	

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

protein - protein search, using sw model

on : December 2, 2001, 16:14:45 ; Search time 48.03 Seconds
(without alignments)
621,704 Million cell updates/sec

title: US-09-072-384-2
perfect score: 2112
sequence: 1 MAGIPGLLFLFLCAVGQ.....IKGNYLDRCREGDTVFLPGSSN 392

scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

searched: 21941 seqs, 76174552 residues

total number of hits satisfying chosen parameters: 219241

minimum DB seq length: 0
maximum DB seq length: 2000000000

post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

database : PIR_68:*

- 1: pir1:*
- 2: pir2:*
- 3: pir3:*
- 4: pir4:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	DB ID	Description
1	185	8.8	A45134	endopeptidase (EC 3.4.21)
2	130.5	6.2	A35122	metalloproteinase
3	125	5.9	482	coagulation factor
4	118	5.6	B26823	pancreatic elastas
5	114	5.4	258	S70439
6	114	5.4	267	4 A56615
7	114	5.4	488	probable Pancreati
8	114	5.4	492	coagulation factor
9	113	5.4	269	pancreatic elastas
10	113	5.4	A26823	pancreatic elastas
11	112.5	5.3	271	trypsin (EC 3.4.21)
12	112	5.3	259	pancreatic elastas
13	112	5.3	266	pancreatic elastas
14	112	5.3	266	hypothetical prote
15	111.5	5.3	522	trypsin-like prote
16	111.5	5.3	238	R-a reactive factor
17	111	5.3	696	complement factor
18	110	5.2	1 ELRT1	pancreatic elastas
19	109	5.2	269	pancreatic elastas
20	107.5	5.1	761	brain-specific ser
21	106	5.0	405	probable secreted
22	106	5.0	236	T-cell suppressor
23	104.5	4.9	271	pancreatic elastas
24	104.5	4.9	273	hypothetical prote
25	104.5	4.9	1582	hypothetical prote
26	102.5	4.9	2055	hypothetical prote
27	102	4.8	786	serine proteinase
28	101.5	4.8	583	complement factor
29	101.5	4.8	747	hypothetical prote

330	101	4.8	1047	2	A55617
331	99	4.7	274	2	S40004
332	97.5	4.6	259	2	S68424
333	97.5	4.6	272	2	JC4170
334	96	4.5	407	1	KFB07
335	94.5	4.5	409	1	T35118
336	94.5	4.5	416	1	KFB0
337	94.5	4.5	1238	2	T34929
338	94.5	4.5	2145	2	JC4747
339	94	4.5	266	2	JC4850
340	94	4.5	624	2	T02289
341	93.5	4.4	452	1	A30351
342	93.5	4.4	548	2	D82175
343	93	4.4	268	2	S68825
344	93	4.4	430	1	A24702
345	92	4.4	260	2	S260A1

ALIGNMENT S

1	A45134	glutamylendopeptidase (EC 3.4.-.-); species: <i>Bacillus licheniformis</i>	glutamate-specific - <i>Bacillus licheniformis</i>
		# Date: 10-Jun-1993 # sequence_revision 18-Nov-1994 #text_change 15-Oct-1999	
		# Accession: A45134; S23078	PID:g21626
		# Cross-references: GB:D100060; NID:g216263; PIDN:BAA00949_1; PID:d1001415; PID:g21626	
		# Experimental source: ATCC 14580	
		# Note: sequence extracted from NCBI backbone (NCBIN:118784, NCBIPI:118785)	
		# Title: Purification, characterization, cloning, and expression of a glutamic acid specific endopeptidase.	
		# Reference number: A45134; Reference number: A45134; MUID:93054737	
		# Accession: A45134	
		# Status: preliminary	
		# Molecule type: DNA	
		# Residues: 1-316 <RKA>	
		# Cross-references: Svedberg, I.; Bredt, K.; Björk, J. Biochem. 204, 165-171, 1992	
		# Title: Isolation and amino acid sequence of a glutamic acid specific endopeptidase.	
		# Reference number: S23078; MUID:9215199	
		# Accession: S23078	
		# Status: preliminary	
		# Molecule type: protein	
		# Residues: 95-116 <SVE>	
		# Keywords: hydrolase	
		Query Match 8.8%; Score 185; DB 2; Length 316;	
		Best Local Similarity 23.2%; Pred. No. 2.9e-08;	
		Matches 77; Conservative 46; Mismatches 139; Indels 70; Gaps 16;	
	63	SSCGPQCHKGRLPL--PTYKEAKQYIQLSYETLYANGSRTPEQVLYILSSSGDGMXXRDSG 119	
	:	: :	
db	28	AQAAAPPH-TPVSSDPSYK-AETSVTYDP-----NIKSDQGYLXSKAFTGTGTG--KVNE 76	
	:	: :	
	120	SSGGKSRRK-----RQIYGDSRSPSIFGKDFLNLAYFSTSVKLST---GCTGTLYAE 167	
	:	: :	
db	77	TKEAKRKSPAKAPYSIKSVIGSDDRTRVTN---TAVPYRAIVHISSIGCTGWMGP 133	
	:	: :	
	168	XHVLTAAHC1HDGKT-YVKGTOQLRGLPKFKEDGRRGANDSTSAMBEQMKFQWIRYKR 226	
	:	: :	
db	134	KTVAGHCLTYDTSSGFAGATVPSG-----RNCTS-----YPIGSVKSTR 175	
	:	: :	
	227	THVKPGWKIGNANDIGMDYDALLEKJPKHKRKFMKIGVSPPAKKQPLPGRHSGYDNDR 286	
	:	: :	
db	176	YFIPSWQRSGNTN---YDGAIELSEPIGNTVGYGYSYTSSLVGTWV1SGYPDK 230	
	:	: :	
	287	PGNLVYRFCD--VKDETDYDLLYQQCDAQPGASGYVYVRMVKRQQQKWERKII---GIF 340	
	:	: :	
db	231	TAGTQOHSGPATAESTYKLOFAM-DTYGGOSGSPVQEQQSSRTINCSPCSLAVHTNSVY 289	

